Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

On Direct vs. Indirect Peer Influence in Large Social Networks

Bin Zhang, Paul A. Pavlou, Ramayya Krishnan

INFORMATION SYSTEMS RESEARCH
Vol. 29, No. 2, June 2018, pp. 292–314
ISSN 1047-7047 (print), ISSN 1526-5536 (online)
http://pubsonline.informs.org/journal/isre/

On Direct vs. Indirect Peer Influence in Large Social Networks

Bin Zhang,[a] **Paul A. Pavlou,**[b] **Ramayya Krishnan**[c]

[a] Eller College of Management, University of Arizona, Tucson, Arizona 85721; [b] Fox School of Business, Temple University, Philadelphia, Pennsylvania 19122; [c] Heinz College, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Contact: binzhang@arizona.edu, http://orcid.org/0000-0003-0675-2222 (BZ); pavlou@temple.edu, http://orcid.org/0000-0002-8830-5727 (PAP); rk2x@cmu.edu, http://orcid.org/0000-0001-9935-2468 (RK)

Received: May 13, 2015
Revised: April 30, 2016; April 10, 2017; August 23, 2017
Accepted: September 27, 2017
Published Online in Articles in Advance: March 2, 2018

https://doi.org/10.1287/isre.2017.0753

Abstract. With the availability of large-scale network data, peer influence in social networks can be more rigorously examined and understood than before. Peer influence can arise from immediate neighbors in the network (formally defined as *cohesion* or direct ties with one-hop neighbors) and from indirect peers who share common neighbors (formally defined as *structural equivalence* or indirect ties with two-hop neighbors). While the literature examined the role of each peer influence (direct or indirect) separately, the study of *both* peer network effects acting simultaneously was ignored, largely because of methodological constraints. This paper attempts to fill this gap by evaluating the simultaneous effect of both direct and indirect peer influences in technology adoption in the context of caller ringback tone (CRBT) in a cellular telephone network, using data from 200 million calls by 1.4 million users. Given that such a large-scale network makes traditional social network analysis intractable, we extract many densely connected and self-contained subpopulations from the network. We find a regularity in these subpopulations in that they consist either of about 200 nodes or about 500 nodes. Using these subpopulations and panel data, we analyze direct and indirect peer influences using a novel autoprobit model with multiple network terms (direct and indirect peer influence, with homophily as a control variable). Our identification strategy relies on a newly designed spatial autoregressive model, allowing us to identify the direct and indirect peer influences on each of the extracted subpopulations. We use meta-analysis to summarize the estimated parameters from all subpopulations. The results show CRBT adoption to be simultaneously determined by *both* direct and indirect peer influence (while controlling for homophily and centrality). Robustness checks show model fit to improve when *both* peer influences are included. The size and direction of the two peer influences, however, differ by group size. Interestingly, indirect peer influence (structural equivalence) plays a negative role in diffusion when group size is about 200, but a positive role when group size is about 500. The role of direct peer influence (cohesion), on the other hand, is always positive, irrespective of group size. Our findings imply that businesses must design different target strategies for large versus small groups: for large groups, businesses should focus on consumers with both multiple one-hop and two-hop neighbors; for small groups, businesses should only focus on consumers with multiple one-hop neighbors.

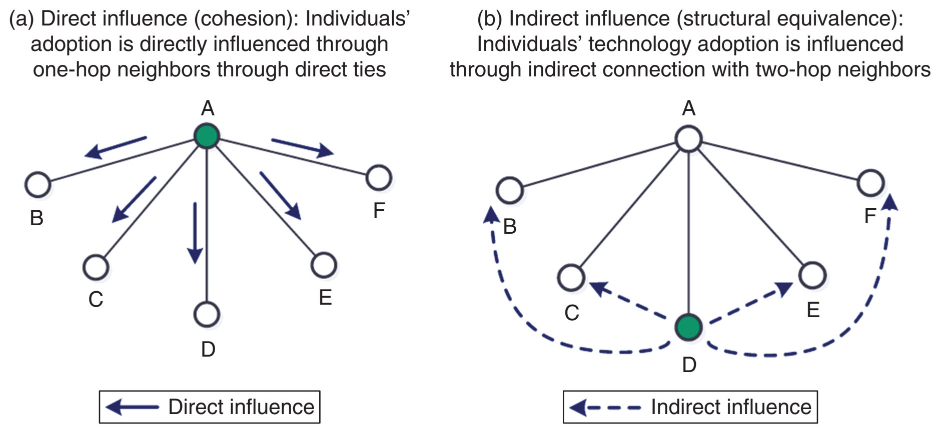
History: Vijay Mookerjee, Senior Editor; Gautam Pant, Associate Editor.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2017.0753.

Keywords: network effects • peer influence • cohesion • structural equivalence • caller ringback tone (CRBT)

1. Introduction

Technology diffusion is a core area of information systems (IS) research. An important assumption made in technology diffusion theory is that adoption is affected by an individual's neighbors in a social network. This process, by which an individual adapts her behavior to those of her neighbors in a social network, is known as *peer influence* (Duncan et al. 1968, Leenders 1997). When making decisions, individuals usually rely on people that they either know directly in their social network (direct ties[1]) or that they know indirectly (indirect ties[2]). Direct peer influence (cohesion) is made through direct communication, that is, direct ties between an individual and her one-hop neighbors (Figure 1(a)); indirect peer influence (structural equivalence) occurs when an individual imitates two-hop neighbors with whom she only has an indirect connection in a social network through others (Figure 1(b)). These two peer influence models lead to very different technology diffusion strategies. For a social network in which diffusion is driven mostly by direct influence, the strategy should be to target individuals with many direct connections, while for a social network driven mostly by indirect influence, the strategy should be to target individuals with many indirect ties. However, while the effect of each of these peer influence models in technology diffusion in social networks has been examined separately, the simultaneous effect of *both*

Figure 1. (Color online) Direct Peer Influence (Cohesion) and Indirect Peer Influence (Structural Equivalence)



peer network influences, especially in large-scale social networks, has not been examined, largely because of methodological challenges. This study aims to fill in this gap by simultaneously examining both direct and indirect peer influences in a large-scale network that makes social network analysis intractable with extant methodological approaches.

The debate about which peer influence—direct (cohesion) or indirect (structural equivalence)—plays a more influential role in technology diffusion is still inconclusive. Coleman et al. (1966) studied diffusion of medical innovation and found that medical doctors adopted new technology at earlier stages largely because of direct influence. However, Burt (1987) reanalyzed Coleman et al.'s data and concluded that diffusion was not driven by direct peer influence, but rather by indirect peer influence. Both camps, those supportive of direct peer influence (e.g., Rogers and Kincaid 1981, Harkola and Greve 1995) and those championing indirect peer influence (e.g., Strang and Tuma 1993, Van den Bulte and Lilien 2001), have found quantitative evidence to support their claims. Thus, determining whether direct influence or indirect influence is the major driver of technology diffusion has been an unresolved research question. This study aims to reconcile these two competing views on technology diffusion in large-scale social networks by examining their effects simultaneously in the activity of cellular phone users in a large network.

Methodologically, we include both direct peer influence and indirect influence as two network terms in a single model. However, most existing analyses only used network autocorrelation models that include one peer influence term at a time or two terms in two separate models. The assumption made by such models is that only one network term can be significant. Some methods were developed to work in situations where both terms are influential, notably Doreian's (1989) two regimes of network effect autocorrelation model and the joint role of both peer influences can be evaluated. However, Doreian's (1989) approach does not support a dichotomous response variable. This is important for technology diffusion where the individual's decision is typically binary—that is, whether to adopt a technology or not. Although Yang and Allenby's (2003) hierarchical Bayes autoregressive mixture model supports a binary response variable, it cannot compare several network terms that may have different signs at the same time. This study introduces a model that not only supports binary response variable and multiple network terms but also permits the signs of multiple network terms to be different. Many IS problems study individuals' binary choices, such as whether to adopt a new technology (or not) or purchase an app (or not). Also, individuals in social networks may be affected by homophily, for which we must also, therefore, control. Such analysis requires a network autocorrelation model having more than two network terms, with homophily as an additional term. However, no model to date can simultaneously examine multiple network terms while still controlling for homophily.

Besides building a model with multiple network terms, we need to solve the data size problem. In terms of network size, our data set contains more than one million nodes (linked individuals). Traditional studies of social networks have been limited to hundreds of nodes. For example, the Coleman et al. (1966) classic medical innovation data only had a node size of 125. Advances in data collection technologies, such as web crawling and Application Programming Interfaces (APIs), have enabled researchers to use large-scale network data. Two major challenges come with big network data. The first and most important is *statistical*. Heterogeneity across subsets of the population increases with the size of the population. As cultures in different parts of the world are distinctive, a network that consists of millions of individuals has groups with various social norms. The pattern of peer influences

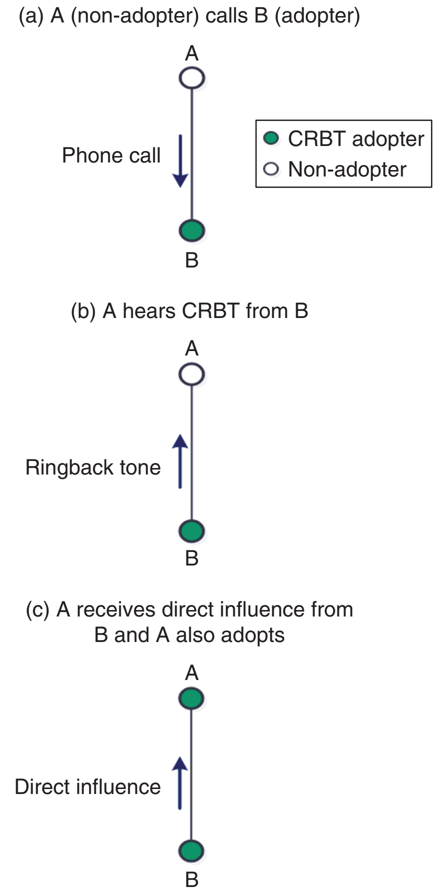
Table 1. Summary of Running Time of Network Analysis Methods/Routines of Various Network Sizes

Analysis method/ routine	Network node size	Network density	Running time
Bonacich (1987) power centrality score	400	0.025	0.37 seconds
	800	0.0125	2 seconds
	1,600	0.00625	14 seconds
	10^6	10^{-5}	> 32 years
Two regimes of network autocorrelation model	400	0.025	9.5 seconds
	800	0.0125	56 seconds
	1,600	0.00625	331 seconds
	10^6	10^{-5}	> 181 years

may be different among these groups. Even assuming that we have sufficient computing power, analysis of the whole population would confound across these distinct effects across different subsets. The second challenge is *computational*. With the exception of a few network analyses, such as degree centrality, most analyses do not scale well. That is, for some class of analytic routines, our standard desktop machines are not fast enough to conduct the analysis of a large network in a realistic time frame. A list of examples run on a standard desktop computer with an Intel 4-core 2.67 GHz CPU and 4 GB RAM is shown in Table 1.

Studying several peer influences in large-scale networks simultaneously presents major methodological challenges. First, comparing peer influence on interdependent decisions among individuals requires a complex statistical model. Accommodating direct and indirect peer influence, together with homophily as a control variable, requires a network autocorrelation model with multiple network terms. However, no model that supports more than two network terms also controls for homophily (Zhang et al. 2013). Second, the analysis of social networks with size at the societal scale (node size > 10^6) cannot be accomplished by existing models. Our approach employs a network processing method capable of handling larger networks, which enables heterogeneity and computation challenges from running the statistical model on the full network to be solved. Comparing peer influences in large-scale networks requires both a new statistical model and a data processing solution for large-scale network data. This study examines three network peer influences in a large social network with a method that can process three network terms in a network with a node size larger than one million (10^6). Our paper makes a methodological contribution by extending the multinetwork autoprobit (mNAP) model (Zhang et al. 2013) to examine the effects of both direct and indirect peer influences on caller ringback tone (CRBT) adoption (Figures 2 and 3), aside from homophily and network centrality, resulting in a novel extension of the mNAP model with simultaneous multiple network terms. To our knowledge, our work introduces the first hierarchical Bayesian model with multiple network

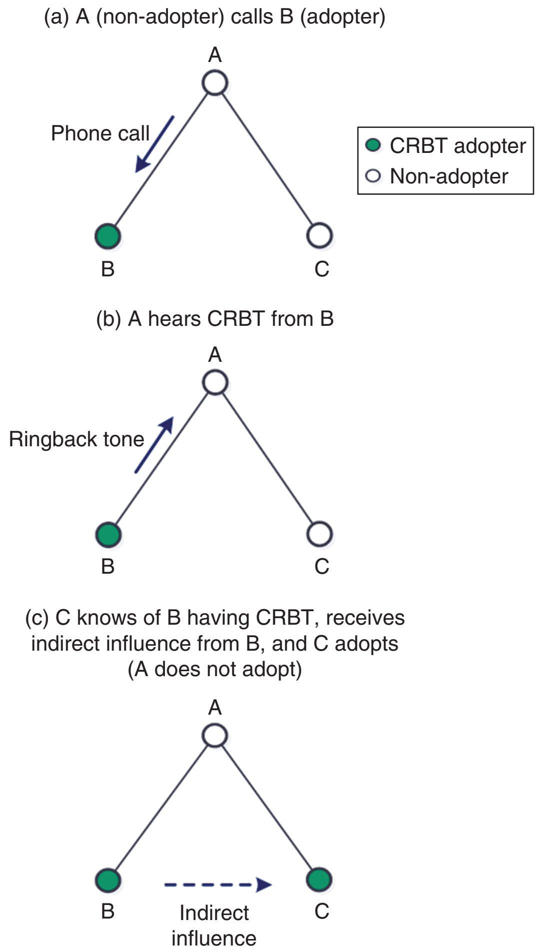
Figure 2. (Color online) Demonstration of Direct Peer Influence of CRBT Adoption (Cohesion)



autocorrelation terms with no restriction on the signs of the estimated parameters, which can be estimated using either cross-sectional or panel data. Finally, following Bramoullé et al. (2009), we demonstrate how peer influences can be identified using our model.

The nature of peer influence, in both magnitude and direction, could be quite different in social networks with a variety of sizes. For example, social networks constructed by friends in small (versus large) social networks may differ, since people in small communities may find it easier to know each other than people in large communities. Therefore, people in small networks are likely to have closer links and have stronger peer influence on each other. Yet, the pattern of the magnitude and direction of peer influences on technology diffusion within different social network sizes, and whether the magnitude and sign of the coefficients of the social network terms change when the size of the social network changes, has, to our knowledge, never been empirically tested. Using large-scale network data, we have the opportunity to examine the role of network size in (direct and indirect) peer influence, thus contributing to the literature on how direct and indirect peer influences vary based on the size of the social network.

Figure 3. (Color online) Demonstration of Indirect Peer Influence of CRBT Adoption (Structural Equivalence)



Our data come from one of the largest cellular phone services in Asia that provide CRBT.[3] There are over a million subscribers and over one billion phone call records in our data (after data preprocessing). Each record includes (hashed and anonymized) phone numbers from the caller and callee, as well as the call date and time. Given the size of our data, it is impossible to analyze the full set using available standard computing power. Thus, to examine CRBT diffusion, we used subpopulations to decrease heterogeneity and also make our analyses computationally tractable. The subpopulations analyzed in our study are extracted using the transitive clustering and pruning (T-CLAP) algorithm (Zhang et al. 2011). People in such dense networks have stronger peer influence than those in sparse networks, and they do not receive contaminated influence from external connections. We developed a new solution to the mNAP model (Zhang et al. 2013) that accommodates three network autocorrelation terms to analyze the data, as the current implementation of mNAP only supports two network terms. Because networks of different sizes could have unique peer influence magnitude or direction, we extracted subnetworks (a densely connected group of individuals) of various sizes. We show that in a large network, when subpopulations (or subnetworks) are extracted using an empirically tested algorithm without specifying a predetermined size, their sizes empirically fall in two levels only—about 200 and 500. These sizes align with the principle of Dunbar's (1992) number, though they are larger, implying that the size of each individual's social network is restricted by her cognitive limit. The number tends to be larger in cellular phone or social networks because of the ease of connectivity.

This study has practical implications, since these two peer influence models will lead to different strategies for businesses. Specifically, businesses could focus on consumers with many direct (one-hop) neighbors or individuals with many indirect (two-hop) neighbors. A cellular phone service provider can easily identify subscribers who have many direct neighbors—those who call and receive calls from many other subscribers—and share direct neighbors with many two-hop neighbors—those who call and receive calls from the same group of people, but never call each other. If CRBT diffusion is driven solely by direct influence, then businesses that want to convert more people to adopt CRBT must target individuals who have many directly connected neighbors in the network, because they can influence a large number of potential adopters. However, if diffusion is driven by indirect influence, then businesses would need to target individuals who have many indirectly connected neighbors, because these individuals will use the indirect neighbors as a reference and thus exhibit the same behavior. If diffusion is driven by both models, a joint scheme needs to be considered. For example, if we want to convert an individual user in the network shown in Figure 1, if the adoption is driven by direct influence, then we need to convert individuals with many ties, such as A, to an adopter, because she will influence all of her direct (one-hop) neighbors. However, if adoption is mainly influenced by indirect peer influence, we need to convert individuals such as C to adopt, because C can affect many individuals, such as B, D, and E, as well as the focal individual, and can eventually influence all of her two-hop neighbors who share the same one-hop neighbors (represented by a dashed arrow in Figure 1(b)). Considering that people are likely to have many two-hop friends, even more than one-hop friends, the effect from indirect influence should not be ignored when evaluating peer influence. To facilitate CRBT adoption in social networks, businesses should target individuals with many one-hop neighbors in smaller groups; in larger groups, businesses should target individuals with both many one-hop neighbors *and* two-hop neighbors.

The rest of the paper is organized as follows. We introduce the literature on direct peer influence and indirect peer influence in the literature review (Section 2). After hypotheses development (Section 3), the statistical model used—mNAP—is described (Section 4). In the data and results (Section 5), we present our data and results. We then explain the key findings and contribution in the discussion (Section 6). Finally, the paper closes with the conclusion (Section 7).

2. Literature Review

2.1. Technology Diffusion in Social Networks

Technology diffusion is the "process by which an innovation is communicated through certain channels over time among the members of a social system" (Rogers 1962, p. 5). The adoption of social media enables us to understand technology diffusion in a large social network (e.g., Brancheau and Wetherbe 1990, Chatterjee and Eliashberg 1990, Premkumar and Nilakanta 1994, Ahuja 2000, Agarwal et al. 2008, Oinas-Kukkonen et al. 2010). Earlier literature assumed that individuals adopt a technology entirely because of their own attributes, such as gender, age, education, and income (Kamakura and Russell 1989, Allenby and Rossi 1998). However, this emphasis on individual attributes may be because of lack of social network data and models for handling large-scale networked data; indeed, studies show that decisions to adopt technology are driven by an individual's social network peers (e.g., Bernheim 1994, Manski 2000, Smith and LeSage 2004). This could be because of a "contagious" effect, where people imitate the behavior of their peers, or a feeling of homophily, in which some unobserved traits drive people's tendency to form a friendship and to exhibit a similar behavior (e.g., Aral et al. 2009, Shalizi and Thomas 2011). Homophily explains how connected individuals with similar attributes make alike decisions (Bott 1928, McPherson and Smith-Lovin 1987, McPherson et al. 2001). In sum, the social contagions literature suggests that social and network aspects affect an individual's technology adoption beyond their individual characteristics.

In a network diffusion model, the process of diffusion is driven by peer influence from neighbors in the network (Valente 2005). Peer influence can be due to *direct influence* (*cohesion*), which is the influence of direct one-hop neighbors, *indirect influence* (*structural equivalence*), influence from two-hop neighbors who share common neighbors, *centrality*, the number of neighbors (Valente 2005), and homophily.[4] Centrality uses the number of direct neighbors to measure the size of influence, without considering the direct neighbors' decisions. Because centrality has been a measure for social influence in the IS literature (e.g., Susarla et al. 2012, Aral and Walker 2011), it is used as a control variable in the model.

Homophily also needs to be controlled for when studying peer influence on interdependent decisions, since homophily may have a confounding effect on such decisions. Although the purpose of our study is not to reconcile peer influence versus homophily on technology diffusion, we still control for homophily when examining the (direct and indirect) effect of peer influence to capture all interdependent adoption decisions.

Several studies have previously considered homophily. Aral et al. (2009) used propensity score matching with observed individual attributes to measure homophily to show that ignoring homophily overestimates peer influence. For example, Nair et al. (2010) used individual-level fixed effects to control for homophily when analyzing peer influence in prescription decisions. Ma et al. (2015) modeled homophily as a group-level correlation, where a group was defined as a clique of size four. In this paper, we define homophily directly from individuals' personal attributes—the weighted average of attribute similarity of all connected individuals. The more two individuals are similar in their demographics, usage behavior, social circle, and location, the higher the homophily between them. Homophily is thus captured by *individual demographics* (age and gender), *location* (frequently used base stations), *behavior* (average call duration in seconds), and *social circle* (number of unique friends called). Taken together, all four dimensions potentially associated with similar adoption decisions among connected individuals, including the three types of peer influence (direct, indirect, and network centrality) other than homophily, are all explicitly accounted for in our model.

In this study, we include both *direct* and *indirect* peer influence as network autocorrelation terms, plus we control for *homophily* and *network centrality*. Centrality, as well as direct and indirect influence, are all derived from the same network in which individuals are embedded. Therefore, our paper makes a contribution to the methodology by including three types of influence when studying technology diffusion. Although the study's main goal is to compare direct and indirect peer influence, we include homophily and centrality in the model as controlled covariates to account for their possible effects on technology diffusion.

2.2. Diffusion Models

In the social networks literature, there are three main models for studying diffusion: macro models, event history model, and network autocorrelation models (Valente 2005).

Early on, the probability of adoption was only related to the time that an actor gets exposed to an object of diffusion. Bass (1969) proposed a model—referred to as the *Bass diffusion model*—that includes both peer influence and innovativeness. This model explains that initial adoption is based on an actor's innovativeness and exposure to potential sources of influence. This model is a population-level model, assuming that every actor in the network has the same probability of interacting with others. We do not use the Bass model, because in large real-world networks, this assumption usually does not hold and the variation of the number of individuals' connections is large (usually following a power law distribution).

Comparatively, the *event history model* offers some useful tools for network analysis. Sometimes, the factors that affect technology diffusion also affect the formation of the network. Therefore, it is necessary to collect data at different time periods (panel data) for the event history analysis. The purpose of event history analysis is to explain why certain people are at a higher risk of experiencing the event of interest than others. The most commonly used analysis methods include failure-time models, survival models, and hazard models (when an event is viewed as a transition from one status to another, such as from nonadoption to adoption). One example of the event history model is as follows (Strang and Tuma 1993):

$$\log(\mathbf{h}(t)) = \alpha + \mathbf{X}\boldsymbol{\beta} + \mathbf{X}_t\boldsymbol{\beta}_t + \rho\mathbf{W}\mathbf{y}_t + \epsilon,$$

where vector $\mathbf{h}(t)$ represents hazard rates of all individuals at time t. Matrix $\mathbf{X}$ is the time-invariant covariates of all individuals, its correspondent coefficient is represented by vector $\boldsymbol{\beta}$. Matrix $\mathbf{X}_t$ is the time-variant covariates, and vector $\boldsymbol{\beta}_t$ is the correspondent coefficient. Term $\mathbf{W}$ is the adjacency matrix describing network connections. Binary vector $\mathbf{y}_t$ represents whether an individual adopted at time t. Scalar ρ is the coefficient of term $\mathbf{W}\mathbf{y}_t$. This model studies whether an individual adopts under both time-invariant and time-variant covariates and requires the collection of connection matrices for each time period, which is a formidable task. Therefore, model implementation is scarce. We do not use this model, since real-life networks are in equilibrium status.

The *network autocorrelation model* completes the trio. It is used to study whether connected individuals tend to have the same behaviors, and it originates from the spatial autoregressive (SAR) model (Ord 1975). The general form of such a model is as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{y} + \epsilon,$$

where $\mathbf{y}$ is an individual's binary decision, where 1 stands for technology adoption of CRBT and 0 otherwise. Vector $\mathbf{X}$ is the exogenous covariates and vector $\boldsymbol{\beta}$ is the correspondent coefficient. Adjacency matrix $\mathbf{W}$ describes the network structure, and scalar ρ is the correspondent coefficient for the autocorrelation term $\mathbf{W}\mathbf{y}$. The mNAP model we use in this study belongs to this category of (network autocorrelation) models.

Most network autocorrelation models can only accommodate one type of peer influence as a single autocorrelation term and compare the coefficients from multiple models using a Q-test (Leenders 2002). Doreian (1989) designed two regimes of a network effect autocorrelation model that only supported a continuous response variable and with only two network impact terms at most. Fujimoto and Valente (2012) directly put network autocorrelation terms on the right-hand side of their logistic regression. Doreian (1982) called this model quick and dirty (QAD), because it did not satisfy the assumption of logistic regression that observations are not independent (and thus the estimation results are biased). Another model under this type, considered state of the art by many, is Yang and Allenby's (2003) hierarchical Bayesian autoregressive mixture model that supports a categorical dependent variable and multiple component networks. Still, this model is not sufficient, because the sum of all weights of component networks must be one, implying that the effect from all network terms must be on the same side and be statistically significant at the same time. Accordingly, to use this family of methods, we need to design our own model.

Another family of social network models are exponential random graph models (ERGM) (Robins et al. 2007) and their extension on dynamic networks temporal ERGM (temporal ERGM; Hanneke et al. 2010). ERGM models assume that the probability of connections among individuals follows an exponential family distribution. However, the objective of these models is predicting network structure using observed network and individual attributes, and they do not align with our focus, which is designed to study interdependent decisions among connected individuals; thus, we are not able to use this family of models in our analysis.

Finally, another popular application for social networks analysis is SIENA (simulation investigation for empirical network analysis) (Snijders and van Duijn 1997, Snijders 2001), which is based on stochastic actor-oriented models and used to represent the evolution of networks. The network structure change is the result of actors (individuals) making changes about network connections based on objective functions. Although extensions of SIENA can model interdependent decisions among connected individuals based on the assimilation principle (Snijders et al. 2010), we could not use SIENA because of two reasons: first, SIENA assumes that individuals can observe the whole network, including all other individuals' attributes and their network connections, and consequently adjust their network connections with their peers. Given large network data (more than one million individuals), such an assumption does not hold in our context. Second, SIENA does not guarantee convergence of the parameter estimation (Steglich et al. 2006). Based on pilot studies, this limitation of SIENA does apply to our data. Therefore, SIENA is inappropriate for our study.

3. Hypotheses Development

3.1. Direct Peer Influence Model

The direct influence (or "cohesion") model, suggests that an individual's technology adoption can be affected by the neighbors to whom she is directly connected. Connections signify communications among individuals in the network. An individual could be informed by, persuaded by, or get suggestions from her neighbors in the network. Coleman et al. (1966) produced perhaps the most famous piece of work that uses a direct peer influence model to explain technology diffusion. They found that medical doctors prescribe a new drug because of direct ties with other doctors. The stronger a doctor's connection to her colleagues (who are also adopters), the earlier the doctor prescribes that drug for the first time. Coleman et al. (1966) gave a very reasonable explanation: when there is a need to make a decision in an ambiguous situation, doctors would ask for suggestions and advice from those with whom they usually discuss questions or from whom they get advice. "The more frequent and empathic the communication is between individual and neighbor" (Burt 1987, p. 1289), the more likely the neighbor's adoption will affect the individual's adoption decision. Such adoption is the result of discussions on benefits and costs between the individual and her neighbors. Rogers and Kincaid (1981) also examined the effect of direct influence (cohesion) on innovation diffusion. By contrast to Coleman et al. (1966), Rogers and Kincaid (1981) used personal network density as a measure of direct influence. Still, their result is similar to that of Coleman et al. (1966)—personal network density is positively related to adoption. Existing IS literature examined peer influence on people's different technology adoption decisions, such as Facebook apps (e.g., Aral and Walker 2011), home computers (e.g., Goolsbee and Klenow et al. 2002), online social networks (e.g., Katona et al. 2011, Trusov et al. 2010), and online content generation (e.g., Susarla et al. 2012, Shriver et al. 2013). The literature also noted that individuals do notice and understand new technologies through discussion and observation with others who are close to their social network (Harkola and Greve 1995), specifically direct neighbors. In the CRBT context, direct influence (cohesion) assumes callers who make phone calls to each other will hear the callee's CRBT and would thus be more likely to become interested in CRBT and eventually buy CRBT. Thus, an individual's direct ties to existing adopters influence her decision to adopt. Therefore, we propose to test whether the average probability of CRBT adoption by people whom an individual calls affects the individuals' probability of CRBT adoption.[5]

We thus propose the following hypothesis for testing:

Hypothesis 1 (H1) (Direct Peer Influence and CRBT Adoption). *An individual's probability of CRBT adoption is positively associated with the number of CRBT adopters among the individual's direct peers* (*one-hop neighbors*).

3.2. Indirect Influence Model

An individual's adoption can also be affected by peers to whom she is indirectly connected, that is, individuals who are friends with a mutual, focal individual (two-hop neighbors) (Figure 1(b)), a relationship that lends itself to an effect termed indirect influence (structural equivalence). Indirect influence is also known as the role equivalence model (Burkhardt 1994).[6] Indirect influence is a measure of the extent to which individuals communicate with the same other people and not necessarily with one another. The degree to which an individual and her peers interact with the same others reflects the extent of their indirect influence. Thus, two individuals may have indirect influence between them, even if they never communicate with one another. An individual would infer judgment of her peers who have the same position on certain things in the influential flow of the social network, and in order not to lose her influential power, the individual would eventually make the same judgment and same decision as well. Indirect influence models "were developed . . . explicitly as a vehicle for describing the structure of role relations defining statuses across multiple networks" (Burt 1987, p. 1293). For example, a medical doctor wants to maintain an image of innovativeness. After another doctor with whom she shares common friends adopts a new technology, the doctor believes the adoption of such technology will enhance her reputation as an innovator and effective power in the social network. Thus, she also wants to adopt. Burt (1987) reanalyzed the Coleman et al. (1966) medical innovation data and concluded that direct influence (cohesion) was not the dominant or the only factor driving diffusion. He found that individuals indirectly connected to and sharing common friends with adopters are more likely to adopt, concluding that the effect of social contagion was through indirect influence instead of direct influence. Strang and Tuma (1993) found weak influence among doctors who are directly connected, but found strong influence between those that are indirectly connected but share common friends. Burkhardt (1994) also compared these two peer influences—direct and indirect—with regard to users' attitudes, self-efficacy, and frequency of using computers. Burkhardt (1994, p. 893) found: "when people evaluate their own personal skills or self-images, they rely on those close to them; when they determine job-related attitudes, they are more likely to rely on structural equivalents." Van den Bulte and Lilien (2001) reanalyzed the Burt (1987) analysis to compare the effects of direct and indirect influence. Distinct from Burt (1987), who used the Euclidean distance to measure indirect influence, Van den Bulte and Lilien (2001) used the proportion of exact peer matches as their measure of indirect influence. Their results showed both direct and indirect influence to be significant, although indirect influence was stronger. Finally, Burt (1987, p. 1293) concluded from the literature that people in a social network will "use each other as a frame of reference for subjective judgments and so make similar judgments even if they have no direct communication with each other."

In the context of CRBT, indirect influence refers to the situation where users in the network have the same or similar pattern of relations with their two-hop neighbors (Figures 2 and 3). In a network constructed by callers and callees, an individual could learn from a direct friend that an indirect neighbor with whom she shares many common neighbors is using CRBT, even though the direct friend may not be a CRBT adopter. The focal user may like the idea of CRBT, as the indirect friend did. The more common friends the focal individual shares with the two-hop friend, the higher the chance the focal user would learn about and thus adopt CRBT. In the indirect influence model, "the trigger to ego's adoption is adoption by the people with whom he jointly occupies a position in the social structure" (Burt 1987, p. 1294). Decision similarity occurs when users adopt the same behaviors as their direct neighbors. Similarity could also happen when a user and her neighbors connect to the same set of people in a CRBT network, that is, when people are indirectly connected to each other (Coleman et al. 1966; Burt 1982, 1987; Krackhardt and Stern 1988). We define indirect influence as the Euclidean distance between two (CRBT) callers, measured by how many common friends two callers share. The more common friends two callers share, the smaller their Euclidean distance is. Thus, we propose the following:

Hypothesis 2 (H2) (Indirect Influence and CRBT Adoption). *An individual's probability of CRBT adoption is positively associated with the number of indirect peers (two-hop neighbors) who share many common direct (one-hop) neighbors who have also adopted CRBT.*

3.3. Network Size

A social network usually consists of clusters, which are subnetworks that are densely connected. As the size of the population gets larger, the heterogeneity across clusters in the population also increases. One source of heterogeneity is the size of the cluster. According to Dunbar's number, the number of stable social relationships an individual can maintain is restricted by her cognitive ability, and varies between 100 and 230 (Dunbar 1992). Therefore, we would believe that in a small social network, the relationships among individuals tend to be stronger; accordingly, the size of the influence will be larger too. Likewise, in a large social network, peer influence among neighbors tends to be smaller. For indirect influence, however, the larger the social network, the more likely an individual has more two-hop neighbors. It is worth noting that the individual's personality may also have an impact on the purchase choice (Ma et al. 2015). This is because CRBT choice can indicate an individual's taste in music. Among indirectly connected neighbors where competition for attention from peers is significant, individuals may want to make a different choice in CRBT to reflect her personality. In larger networks, the competition for attention is typically more intense, and the attempt to showcase one's personality is expected to be higher than smaller networks. Thus, our hypotheses about the size of the network on both network peer influence terms are proposed as follows:

Hypothesis 3 (H3A) (Network Size and Direct Influence). *The effect of direct influence (cohesion) on an individual's probability of CRBT adoption is negatively associated with the size of the network to which the individual belongs.*

Hypothesis 3 (H3B) (Network Size and Indirect Influence). *The effect of indirect influence (structural equivalence) on an individual's probability of CRBT adoption is positively associated with the size of the network to which the individual belongs.*

4. Method

4.1. Model

The goal of our study requires the simultaneous accommodation of two network autocorrelation terms (direct and indirect peer influence), allowing these two terms to have different signs without having to be significant at the same time. In addition, a third network autocorrelation term is included to account for homophily. The mNAP model (Zhang et al. 2013) becomes a natural choice because of its support for a binary outcome, covariates describing individuals in terms of their network attributes, and most importantly, more than two network autocorrelation terms. However, only an analytical solution for including more than two network autocorrelation terms is provided for mNAP. A solution for mNAP with multiple network autocorrelation terms explicitly controlling for homophily has never been implemented. Therefore, we not only offer an explicit specification of the mNAP model (Zhang et al. 2013) with two network autocorrelation terms to study both direct and indirect peer influence on CRBT diffusion, we include two terms for homophily from direct and indirect friends, respectively. We modeled peer influence as latent and correlated terms among individuals. Following Aral et al. (2009), we used observed individual attributes to measure homophily, as described below. Centrality, another type of peer influence is also controlled for, but as an observed covariate. Notably, the proposed model supports longitudinal panel data, making it the first, to our knowledge, hierarchical Bayesian model to take homophily into consideration. With all these factors being included, the mNAP model takes both the interdependence of the individuals' decision, direct peer influence, indirect peer influence, and homophily, and individuals' observable attributes, such as centrality, into consideration. Taken together, the specification of our model, extending mNAP, is described as[7]

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{1}(\mathbf{z}_t > 0),\\
\mathbf{z}_t &= \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\theta}_t + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_t,\\
\boldsymbol{\theta}_t &= \rho_1\mathbf{W}_{1,t}\boldsymbol{\theta}_t + \rho_2\mathbf{W}_{2,t}\boldsymbol{\theta}_t + \boldsymbol{\rho}_3\mathbf{W}_{1,t}\mathbf{H}_t + \boldsymbol{\rho}_4\mathbf{W}_{2,t}\mathbf{H}_t + \mathbf{u}_t,\\
\boldsymbol{\varepsilon}_t &\sim \text{Normal}_n(0, I_n),\\
\mathbf{u}_t &\sim \text{Normal}_n(0, \sigma^2 I_n),\\
\mathbf{X}_t &= \boldsymbol{\mu} + \mathbf{e}_t,\\
(\boldsymbol{\eta}, \mathbf{e}_t) &\sim \text{MVN}(0, \boldsymbol{\Omega}),
\end{aligned}$$

where $\mathbf{y}_t$ is the vector of observed binary choices—whether a caller purchased CRBT—in time period t. Function $1(\cdot)$ represents an indicator function of the latent preference of users, $\mathbf{z}_t$. If $\mathbf{z}_t$ is larger than a threshold 0, users choose $\mathbf{y}_t$ as 1—purchase CRBT; if $\mathbf{z}_t$ is smaller than 0, then users would choose $\mathbf{y}_t$ as 0, that is, not purchase CRBT. The latent preference vector $\mathbf{z}_t$ could be represented as a function of exogenous covariates $\mathbf{x}_t$, autocorrelation term $\boldsymbol{\theta}_t$, and fixed effects $\boldsymbol{\eta}$. The covariates that make up vector $\mathbf{X}_t$ are exogenous degree centrality of the cellular phone subscribers, which is measured by the number of unique callees to which a focal subscriber places phone calls. Matrix $\mathbf{X}_t$ consists of two parts, a time-invariant mean $\boldsymbol{\mu}$ and a time-variant error term $\mathbf{e}_t$. We considered adding both one-hop and two-hop neighbor percentages. Our concern was that these two percentages are highly correlated with degree centrality and may not be necessary to include. First, the percentage of one-hop neighbors is d_i/N, where d_i is the degree centrality of individual i and N is the network size. Since N is a constant for all individuals in the same network, the percentage d_i/N is equivalent to degree centrality d_i. Second, the percentage of indirect friends is given by $\sum_j d_j/N$, where j is the neighbor of the focal individual. Since the aggregation of each individual's degree centrality is already controlled for, such a measure may be redundant. Our concern is supported empirically; when including both percentage variables in our model, neither of them are significant.

Covariance of η and $\mathbf{e}_t$ follows a multivariate normal distribution. Vector $\boldsymbol{\theta}_t$ is the autocorrelation term, which is responsible for the nonzero covariances in $\mathbf{z}_t$. It can be described as the sum of the product between network structures and unobserved preference $\mathbf{W}_{i,t}\boldsymbol{\theta}_t$. Matrix $\mathbf{W}_{i,t}$ represents the social structure underlying each autocorrelation term. Matrix $\mathbf{W}_{1,t}$ describes connections between one-hop neighbors, which is the normalized adjacency matrix, in time period t; matrix $\mathbf{W}_{2,t}$ describes two-hop neighbors in time period t. The definition of all $\mathbf{W}_{i,t}$ is given below. Scalars ρ_1 and ρ_2 are the parameters of two network influence terms, respectively. Considering that the variation of degree (number of one-hop neighbors) is large, we normalized the adjacency matrix when defining $\mathbf{W}_{1,t}$ as

$$\mathbf{W}_{1,t} = \left\{\frac{a_{ij,t}}{\sum_j a_{ij,t}}\right\},$$

where $a_{ij,t}$ is the entry in the adjacency matrix representing individual connections in time t, $\mathbf{A}_t$, $\mathbf{A}_t = \{a_{ij,t}\}$.

The definition of indirect connections is based on indirect peer influence. We followed the literature on the comparison between direct and indirect peer influence done by Van den Bulte and Lilien (2001), and we used the Euclidean distance between two individuals as the measure for indirect connection, which is a type of social position similarity. In an undirected network with unweighted edges, the Euclidean distance $d_{ij,t}$ between two individuals i and j in time t is the sum of the squared difference between the adjacency vectors of nodes i and j

$$d_{ij,t} = \sqrt{\sum_{k=1,k\neq i,j}^{n} (A_{ik} - A_{jk})^2}, \qquad (1)$$

where $A_{ik,t}$ is 1, if node i and k are neighbors, and 0 otherwise.

Since a larger Euclidean distance represents a lower influence between nodes i and j, we used the inverse of the summation of $d_{ij,t}$ and a small constant—one. Thus matrix $\mathbf{W}_{2,t}$ is defined as

$$\mathbf{W}_{2,t} = \{s_{i,tj}\} = \left\{\frac{1}{d_{ij,t}+1}\right\}.$$

Fang et al. (2013) also used the same measure. $d_{ij,t}$ is defined in Equation (1). The element of $\mathbf{W}_{2,t}$, $s_{ij,t}$, has a positive relationship with indirect influence, where a large value represents a higher indirect influence.

It is worth noting that both the Euclidean distance and cosine similarity are the two most common measures for position similarity. The cosine similarity of nodes i and j is defined as the number of common neighbors divided by the geometric mean of the two nodes' degrees (Salton 1989). Cosine similarity and Euclidean distance are related and can be converted to each other under certain conditions. The squared Euclidean distance is actually proportional to the cosine distance. We present the results using cosine similarity in Online Appendix C (Table C7). From these results, we show there is no significant difference between the models using these two measures of position similarity.

We also defined a mathematical rule to enforce $\mathbf{W}_{2,t}$ not correlated with $\mathbf{W}_{1,t}$. Use operator $\circ$ to represent element-wise multiplication, and define $\bar{\mathbf{A}}$ as a matrix whose elements are the logical NOT of the correspondent element in matrix $\mathbf{A}$. $\mathbf{A}$ is the adjacency matrix used to define $\mathbf{W}_{1,t}$. In simple words, $\bar{\mathbf{A}}$ is defined by switching $\mathbf{A}$'s 1 entry to 0, and 0 entry to 1. Then for all $\mathbf{W}_{2,t}$, we convert them to $\mathbf{W}'_2 = \bar{\mathbf{A}} \circ \mathbf{W}_{2,t}$. All the $\mathbf{W}_{2,t}$ used in the real analysis are the converted $\mathbf{W}'_{2,t}$ that do not have correlation with $\mathbf{W}_{1,t}$.

Term $\mathbf{H}_t$ is the matrix of individual attributes representing different dimensions of homophily in period t. According to the literature, homophily is associated with both time-invariant attributes, such as age, gender, race, and education (McPherson et al. 2001), and time-variant attributes, which affect how individuals behave similarly (Lazarsfeld and Merton 1954). Following the literature, we included multiple attributes representing various dimensions of homophily, including demographics (age and gender). Since homophily may also be associated with behavioral attributes about patterns of people's cellular phone use, we included time-variant homophily dimensions, such as location (frequently used stations), behavior (average call duration (seconds)), and social circle (number of unique friends called). We define $\mathbf{H}_t$ in both product terms $\mathbf{W}_{1,t}, \mathbf{H}_t$ and $\mathbf{W}_{2,t}, \mathbf{H}_t$, respectively, to control for the effects of homophily that may come from both direct and indirect friends. The terms $\boldsymbol{\rho}_3$ and $\boldsymbol{\rho}_4$ are the two correspondent vectors of parameter, respectively.

4.2. Identification Strategy

The major challenge in identifying peer influence is the "reflection problem" (Manski 1993, p. 532)—endogeneity in peer influence, or the inability to separate the endogenous effect (latent peer influence) from the exogenous effect (similarity in individual attributes that may result in homophily). Bramoullé et al. (2009) proposed a spatial autoregressive model with network autocorrelation terms that can be categorized as a network autocorrelation model. They also showed that if a network contains an individual's three-hop neighbors, peer influence is identifiable. The principle of the Bramoullé et al. (2009) strategy is that the behavior of a neighbor's neighbor can be used as an instrumental variable, because it is not a determinant of response variable $\mathbf{y}$, despite still being correlated with individual attributes. Considering a simplified, yet still containing all necessary terms, structural model about $\boldsymbol{\theta}_t$, the core variable representing latent preference of CRBT purchase would be:

$$\boldsymbol{\theta}_t = \rho_1 \mathbf{W}_{1,t}\boldsymbol{\theta}_t + \rho_2 \mathbf{W}_{2,t}\boldsymbol{\theta}_t + \boldsymbol{\rho}_3 \mathbf{W}_{1,t}\mathbf{H}_t + \boldsymbol{\rho}_4 \mathbf{W}_{2,t}\mathbf{H}_t + \mathbf{u}_t.$$

The identification of peer influences is achieved if and only if all of the parameters in the structural model ρ_1 to ρ_4 can be uniquely recovered from the reduced form below

$$\begin{aligned}
&(I - \rho_1\mathbf{W}_{1,t} - \rho_2\mathbf{W}_{2,t})\boldsymbol{\theta}_t = \boldsymbol{\rho}_3\mathbf{W}_{1,t}\mathbf{H}_t + \boldsymbol{\rho}_4\mathbf{W}_{2,t}\mathbf{H}_t + \mathbf{u}_t,\\
&\boldsymbol{\theta}_t = (I - \rho_1\mathbf{W}_{1,t} - \rho_2\mathbf{W}_{2,t})^{-1}(\boldsymbol{\rho}_3\mathbf{W}_{1,t} + \boldsymbol{\rho}_4\mathbf{W}_{2,t})\mathbf{H}_t\\
&\qquad + (I - \rho_1\mathbf{W}_{1,t} - \rho_2\mathbf{W}_{2,t})^{-1}\mathbf{u}_t.
\end{aligned}$$

The model above can be interpreted as instrumental variables (IVs). From the series expansion of the model

$$(I - \mathbf{A})^{-1} = \sum_{k=0}^{\infty} \mathbf{A}^k, \quad \text{where: } \mathbf{A} = \rho_1\mathbf{W}_{1,t} + \rho_2\mathbf{W}_{2,t}.$$

A within network transformation (demean) on $\boldsymbol{\theta}_t$ is given by

$$\begin{aligned}
(I - \mathbf{W}_{1,t} - \mathbf{W}_{2,t})\boldsymbol{\theta}_t &= (I - \mathbf{W}_{1,t} - \mathbf{W}_{2,t})(\rho_1\mathbf{W}_{1,t} + \rho_2\mathbf{W}_{2,t})\boldsymbol{\theta}_t\\
&\quad + (I - \mathbf{W}_{1,t} - \mathbf{W}_{2,t})(\boldsymbol{\rho}_3\mathbf{W}_{1,t} + \boldsymbol{\rho}_4\mathbf{W}_{2,t})\\
&\quad \cdot \mathbf{H}_t + (I - \mathbf{W}_{1,t} - \mathbf{W}_{2,t})\mathbf{u}_t.
\end{aligned}$$

Its reduced form is given by

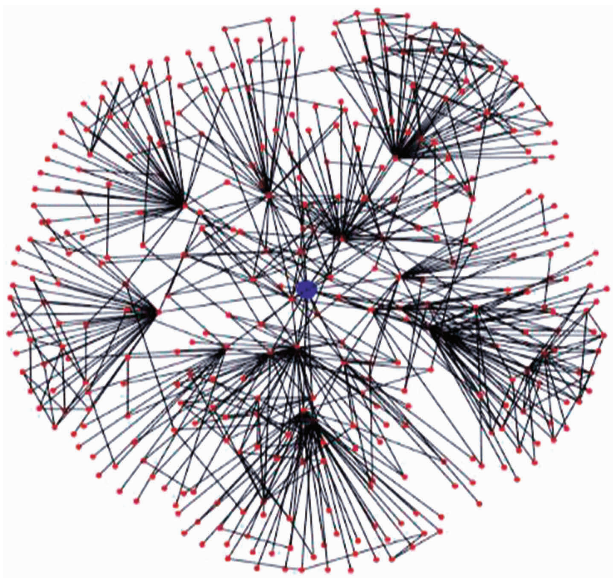
$$\begin{aligned}
&(I - \mathbf{W}_{1,t} - \mathbf{W}_{2,t})\boldsymbol{\theta}_t\\
&\quad = (I - \rho_1\mathbf{W}_{1,t} - \rho_2\mathbf{W}_{2,t})^{-1}(\boldsymbol{\rho}_3\mathbf{W}_{1,t} + \boldsymbol{\rho}_4\mathbf{W}_{2,t})\\
&\qquad \cdot (I - \mathbf{W}_{1,t} - \mathbf{W}_{2,t})\mathbf{H}_t + (I - \rho_1\mathbf{W}_{1,t} - \rho_2\mathbf{W}_{2,t})^{-1}\\
&\qquad \cdot (I - \mathbf{W}_{1,t} - \mathbf{W}_{2,t})\mathbf{u}_t.
\end{aligned}$$

Expected peer influence in the network, conditional on homophily is $\mathrm{E}[(I - \mathbf{W}_{1,t} - \mathbf{W}_{2,t})(\rho_1\mathbf{W}_{1,t} + \rho_2\mathbf{W}_{2,t}) \cdot \boldsymbol{\theta}_t \mid \mathbf{H}_t]$. Using the Bramoullé et al. (2009) proposition, variables $[(I - \mathbf{W}_{1,t} - \mathbf{W}_{2,t})(\boldsymbol{\rho}_3\mathbf{W}_{1,t} + \boldsymbol{\rho}_4\mathbf{W}_{2,t})\mathbf{H}_t^2, (I - \mathbf{W}_{1,t} - \mathbf{W}_{2,t})(\boldsymbol{\rho}_3\mathbf{W}_{1,t} + \boldsymbol{\rho}_4\mathbf{W}_{2,t})\mathbf{H}_t^3, \ldots]$ can be used as valid IVs for $(I - \mathbf{W}_{1,t} - \mathbf{W}_{2,t})(\rho_1\mathbf{W}_{1,t} + \rho_2\mathbf{W}_{2,t})\boldsymbol{\theta}_t$. This condition is equivalent to $I, (\mathbf{W}_{1,t} + \mathbf{W}_{2,t})^2, (\mathbf{W}_{1,t} + \mathbf{W}_{2,t})^3$ are linearly independent, which ultimately can be converted to the condition of both $\mathbf{W}_1$ and $\mathbf{W}_2$'s diameter being greater than or equal to 3. Figure 4 shows an example of an extracted subnetwork of size 252. An edge from node i to node j represents that individuals i and j make mutual phone calls in that period. From the network topology, we can easily find that there are paths between nodes with a distance greater than or equal to three. The linear independence among matrices representing direct and indirect peer influence $I, \mathbf{W}_{k,t}, \mathbf{W}_{k,t}^2$, and $\mathbf{W}_{k,t}^3$ ($k = 1, 2$; 1 = direct influence; 2 = indirect influence) is quantitatively confirmed for each subnetwork extracted in all time periods. Thus, the necessary and sufficient condition of peer influence identification by Bramoullé et al. (2009) is satisfied, as illustrated in Online Appendix B.

5. Data and Results

Our data were obtained from a large Asian telecommunications firm. Data include cellular phone call records and CRBT purchase records over a three-month period, in addition to demographic information, such as age and gender. Since the phone call conversation network is directed, asymmetry can exist between callers. We restricted the data to reciprocal calls, since symmetric connections imply equal and stable connections, while an asymmetric connection indicates an unstable relationship (e.g., Hanneman and Riddle 2005).

Figure 4. (Color online) Example Subnetwork of Size 252 in Time t



We define reciprocity for dyads (A, B) as the condition in which A calls B and B calls A in the same month. We interpreted reciprocity as a higher probability that the two parties are acquaintances. Thus, we further constrained our analysis to include data with reciprocal dyads. Constrained by these requirements, the total size of our phone call record was about 197 million calls placed by about 1.4 million users. This network is not only too large to analyze, but it also contains many clusters that contain different effect sizes in them, and hence it could be analyzed only by using multiple subpopulations and meta-analyses.

A detailed description of the preprocessed data is listed in Table 2. The dependent variable was measured as a binary variable, indicating whether a caller purchased CRBT in a three-month period. We included the degree of the caller and the number of unique callees a subscriber called (and vice versa) to observe the exogenous effect of the number of neighbors. Direct peer influence was captured as callers who make phone calls to each other (0 or 1). Since the number of people callers call are dramatically different across subscribers, we normalized the direct influence matrix by dividing each row by the total number of neighbors, to make the matrix element be the percentage of adopters among their neighbors.

Indirect peer influence is herein defined as the inverse of the Euclidean distance of the neighbor list (adjacency vector) of callers. Indirect influence is only calculated among callers who are not directly connected to guarantee that direct and indirect peer influences are not correlated. The matrix $\mathbf{H}_t$ of individuals' attributes representing different dimensions of homophily in time period t, including the multiple

Table 2. Description of Variables

Variable	Description
$\mathbf{Y}_t$	Dependent variable, whether subscriber purchase CRBT (binary) in time period t
Gender	Gender of cellular phone account holder
Age	Reported age of cellular phone account holder
Degree centrality$_t$	Number of callees that a subscriber has called in time period t
$\mathbf{W}_{1,t}$	Matrix describing direct connection in time period t, normalized adjacency matrix
$\mathbf{W}_{2,t}$	Matrix describing indirect connection time period t, elements are defined by inverse of Euclidean distance of adjacency vector
$\mathbf{H}_t$	Matrix describing homophily, each column represents one measure. Measures include dimensions about homophily such as demographics, location, behavior, and social circle.

dimensions of homophily defined by McPherson et al. (2001). Attributes used to control for homophily are individual demography (age and gender), location (frequently used base stations), behavioral patterns (average call duration, in seconds), and social circle (number of unique friends called). Accordingly, variables included in the $\mathbf{H}_t$ matrix were age_i and $gender_i$, representing the age and gender of individual i, respectively; $duration_{i,t}$, the average call duration in seconds of i in time period t; $uniqfrnd_{i,t}$, the number of unique friends of i in time t; and $location_{ik}$, the most frequent location of individual i in time t.

5.1. Data

Before examining the role of network peer influence in CRBT adoption, we faced another challenge: the size of the data. As a social network with 1.4 million users, our network of focus is quite large. The size of the data introduces two challenges—one *statistical* and one *computational*. The *statistical* challenge is such that: as the size of the network grows, the heterogeneity across subsets of the population also increases. The *computational* challenge manifested itself in most social network analysis packages not scaling well. That is, for some class of questions and analytic routines, standard desktop systems are not able to analyze large networks in a realistic time frame. Memory-wise desktops do not have enough power to accommodate the structure of such a large social network. One solution to this problem is using subpopulations of a smaller size that is computable within the restrictions of memory size in a realistic time frame. We wanted subpopulations to have the following favorable characteristics: first, they should have high density within the network. Internal density shows strong connections among individuals in a network, so social influence is more likely to happen. Second, these networks should still have variation in their connections. Third, there should be relatively few ties from within the subpopulation to the total network. We wanted to avoid "boundary leakage" (nodes with more edges to the external network than the internal network that contaminate the structure of the extracted networks). Fourth, all these subpopulations should have relatively small node sizes so the estimation of the model could be finished in a reasonable amount of time. Since no algorithm exists to provide an ideal balance of quality and speed, we used the T-CLAP algorithm (Zhang et al. 2011) to identify dense and relatively independent subpopulations. T-CLAP outperformed leading algorithms in community detection, such as Infomap, which can also be used for subpopulation extraction. The algorithm does not require returned network size as a parameter, and thus it does not predetermine the subpopulation size. Instead, it returns a subpopulation with a local maximal *I-E* ratio, which is a measure for cluster quality and defined as the ratio of the total internal degree to total external degree of a network. The algorithm uses the *I-E* ratio to define the boundaries among individual groups. Each individual is placed within the group (subnetwork) from which she has a connection. If an individual has connections with more than one group, she is placed in the group with which she has the most connections. Thus, using this principle, each individual is placed in a subnetwork from which she receives the most peer influence. As a result, we avoid contaminated influence from external networks. The size of the returned subnetwork is also close to the true group size; this is because subnetworks representing individual groups are well reflected by the design of the *I-E* ratio—densely connected internally and relatively sparsely connected with external networks. Note that the T-CLAP algorithm does not exclude weak ties. We extracted more than 100 subpopulations, covering more than 50% of the nodes in the entire population. Interestingly, the size of the returned subpopulation can only be sorted into two categories, one of approximate size 200 and the other of approximately 500. These two numbers only represent the approximate magnitude of the networks, not their exact size. For ease of reference, we define the group of about 200 as the "smaller group," and we define the group of about 500 as the "larger group." All estimation results generally follow the same pattern in each size level, so we randomly picked five from the ≈200 level and five from the ≈500 level. This pattern also suggests that in the cellular phone call social network, the number of contacts an individual could manage is larger than Dunbar's (1992) number. The size of subnetworks, *I-E* ratio, where a higher value indicates denser and more cohesive internal connections, and the density of all 10 subpopulations is listed in Table 3.

Descriptive statistics of the independent variables for each subpopulation are listed in Table 4.

Table 3. Extracted Subpopulations' Structure Characteristics

	Subpopulation	n	I-E ratio	Density
Smaller group	1	191	0.83	0.13
	2	202	0.43	0.10
	3	213	0.48	0.039
	4	238	0.82	0.030
	5	263	0.82	0.029
Larger group	6	465	0.43	0.075
	7	485	0.82	0.033
	8	553	0.82	0.029
	9	563	0.43	0.021
	10	597	0.82	0.082

Table 5 presents the results of model estimation for all 10 subpopulations whose node size ranges from 191 to 597. We found that direct influence (cohesion) is consistent across all subpopulations. The results show that direct peer influence from all subpopulations is significant at the $p < 0.05$ level, ranging from 0.056 to 0.075, thus supporting H1. It shows that callers receive strong peer influence through direct connections to one-hop neighbors in the same group who have already adopted CRBT. The simple explanation is that if a caller calls more CRBT subscribers, she gets exposure to more ringback tones, and she is more likely to hear ring tones that interest her, which means that she is more likely to buy those ring tones. Finally, the effects of centrality (degree) are positive and significant at the $p < 0.01$ level, which suggest that if a caller calls more people, the probability of adopting CRBT increases. This result is also consistent with the support for H1.

We also observed a significant effect of indirect peer influence, which suggests that the adoption of CRBT is also impacted by two-hop neighbors. For the indirect influence model, a caller evaluates the behavior of two-hop neighbors who are in the same position of a phone call network as she is (Figure 3). Being in the same position in a social network means that people are in similar relationships with the same group of people. Interestingly, the effect size of indirect influence varies with the size of the subpopulation. When the subpopulations consist of about 200 individuals, the effect of indirect influence is significant but negative, meaning that individuals with more CRBT adopters who have the same social position as they have in the network (two-hop neighbors) are less likely to adopt CRBT. This finding rejects H2, and while it seems surprising at first glance, it is actually quite reasonable. One explanation is that when adopting CRBT, people not only want to be fashionable, but they also want to reflect their individuality. In a smaller group, people weigh individuality higher than being fashionable or making similar decisions. If two-hop friends adopt CRBT, the focal individual does not want to do the same thing but rather to show her individuality by having a different adoption decision

Table 4. Descriptive Statistics of Independent Variables

	Subpopulations									
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
N	191	202	213	238	263	465	485	553	563	597
Gender	0.19	0.12	0.033	0.11	0.065	0.57	0.12	0.27	0.12	0.050
(0 = *male*)	(0.50)	(0.45)	(0.18)	(0.40)	(0.29)	(0.89)	(0.41)	(0.62)	(0.42)	(0.25)
Age	45	42	38	42	42	48	45	40	42	36
	(11)	(12)	(11)	(11)	(12)	(12)	(13)	(13)	(13)	(12)
Centrality	13.6	10.0	3.6	4.0	4.5	18.1	9.3	9.8	5.3	3.4
	(14.9)	(10.6)	(5.1)	(5.2)	(5.6)	(18.6)	(13.6)	(15.5)	(6.7)	(1.8)
Location	0.15	0.13	0.17	0.13	0.27	0.13	0.16	0.29	0.24	0.17
	(0.39)	(0.24)	(0.13)	(0.40)	(0.30)	(0.51)	(0.24)	(0.39)	(0.41)	(0.19)
Behavior	107.7	158.6	127.0	180.2	178.5	112.3	102.6	120.8	132.3	98.6
	(53.5)	(28.4)	(62.2)	(40.0)	(35.5)	(22.0)	(66.3)	(63.5)	(49.4)	(66.5)
Social Circle	22	16	9	7	7	33	21	19	11	5
	(24)	(6)	(9)	(12)	(11)	(38)	(24)	(33)	(8)	(3)

Note. The numbers in the top row of each cell are the means, and the numbers in parentheses are the standard deviations.

(not to adopt). When the subpopulations are larger (about 500), however, indirect peer influence is significant and positive, thus supporting H2. Considering the centrality of bridge nodes connecting indirect neighbors may account for such difference, we empirically tested the difference between the centralities of bridge

Table 5. Results of Analysis Using the mNAP Model (Direct vs. Indirect Influence)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Subpopulations					
Degree	0.014**	0.040*	0.030*	0.0043	0.069*	0.040*	0.042*	0.023	0.043*	0.024*
	(0.0088)	(0.0098)	(0.0095)	(0.020)	(0.020)	(0.022)	(0.019)	(0.033)	(0.044)	(0.0070)
Direct Influence	0.070*	0.060*	0.065**	0.056**	0.061*	0.016*	0.013*	0.014*	0.016**	0.014*
	(0.022)	(0.022)	(0.026)	(0.027)	(0.019)	(0.0048)	(0.0045)	(0.0045)	(0.0063)	(0.0042)
Indirect Influence	−0.0059***	−0.0098**	−0.010**	−0.0079**	−0.0015**	0.0066**	0.0098**	0.0010**	0.034**	0.0049**
	(0.0031)	(0.0041)	(0.0045)	(0.0035)	(0.00080)	(0.0030)	(0.0041)	(0.00060)	(0.015)	(0.0024)
					Direct Homophily					
Gender	0.38	−0.20	−0.034	0.27	0.20	−0.11	0.14	0.69	0.094	0.90
	(0.51)	(3.3)	(0.33)	(0.51)	(0.22)	(0.31)	(0.16)	(0.42)	(0.46)	(0.87)
Age	0.023	−0.026	0.053	0.024	0.028	0.0086	0.064	0.037	0.065	−0.080
	(0.018)	(0.017)	(0.044)	(0.025)	(0.027)	(0.021)	(0.054)	(0.033)	(0.040)	(0.067)
Location	0.016	0.018	0.0050	0.017	0.024	0.017	0.021	0.0061	−0.0071	0.022
	(0.012)	(0.019)	(0.0070)	(0.014)	(0.017)	(0.016)	(0.017)	(0.0055)	(0.0077)	(0.011)
Behavior	0.026	0.044	0.0091	0.028	0.045	0.039	0.040	0.011	0.013	0.053
	(0.021)	(0.031)	(0.014)	(0.024)	(0.043)	(0.19)	(0.028)	(0.0090)	(0.021)	(0.041)
Social Circle	0.055	0.050	0.014	0.045	0.061	0.048	0.058	0.018	−0.020	0.065
	(0.033)	(0.035)	(0.011)	(0.029)	(0.044)	(0.031)	(0.037)	(0.011)	(0.013)	(0.051)
					Indirect Homophily					
Gender	0.036	−0.018	−0.003	0.019	0.014	−0.008	0.010	0.058	0.0080	0.078
	(0.045)	(0.029)	(0.028)	(0.040)	(0.017)	(0.0080)	(0.011)	(0.033)	(0.040)	(0.083)
Age	0.0018	−0.0025	0.0050	0.0025	0.0023	0.00077	0.0049	0.0033	0.0042	−0.0069
	(0.0013)	(0.0015)	(0.0037)	(0.0020)	(0.0019)	(0.0020)	(0.0038)	(0.0028)	(0.0030)	(0.0054)
Location	0.0015	0.0012	0.00038	0.0012	0.0023	0.0016	0.0015	0.00042	−0.00051	0.0016
	(0.00098)	(0.0018)	(0.00062)	(0.0011)	(0.0013)	(0.0015)	(0.0012)	(0.00051)	(0.00052)	(0.0011)
Behavior	0.0019	0.0032	0.00070	0.0021	0.0035	0.0030	0.0030	0.00078	0.00088	0.0041
	(0.0018)	(0.0022)	(0.00094)	(0.0024)	(0.0034)	0.018	0.0020	0.000091	0.0015	0.0037
Social Circle	0.0040	0.0037	0.0011	0.0034	0.0047	0.0036	0.0043	0.0013	0.0013	0.0050
	(0.0029)	(0.0025)	(0.00074)	(0.0029)	(0.0035)	(0.0030)	(0.0026)	(0.0011)	(0.00091)	(0.0054)

$^{*}p < 0.01$; $^{**}p < 0.05$; $^{***}p < 0.10$.

nodes in smaller and larger networks, respectively, and there is no significant difference. Taking both results above about smaller networks ($n \approx 200$) and larger networks ($n \approx 500$), we observe a positive association between indirect peer influence and network size, confirming H3B. One explanation is that in a phone social network, people who call each other are likely to be acquaintances by belonging to the same social group (under some indirect relationships). The enthusiasm of demonstrating to others her adoption of a fashionable product is high, and the satisfaction of letting friends appreciate one's fashion statement is thus also high. Motivated by this thought, the individual knows that her two-hop neighbors whom she does not necessary call have adopted CRBT through common friends they both call, and she is likely to imitate the indirect neighbor. The more ring tones those neighbors bought, the more CRBT the focal person is also likely to buy. In fact, individuals in larger networks can still reflect their individuality even when making the same adoption decision by purchasing CRBT of a different genre. To test this explanation, we empirically examined the genres of CRBT purchases due to indirect peer influence. Controlling for all observables in our data set, in larger networks, the majority of adopters (80% on average) under indirect peer influence purchase CRBT of a different genre than those of their indirect neighbors. This ratio is only about 50% in smaller networks. This shows that if her indirect neighbors adopt CRBT, the focal node wants to be different and thus not adopt. Since the results imply that network size may be correlated with subpopulation size, a meta-analysis that integrates the estimated parameters for direct and indirect peer influence based on subpopulation size was performed. Specifically, the coefficient of direct influence has a higher significance level and size than that of indirect influence, implying that direct peer influence is more influential than indirect influence on CRBT adoption. Peer influence has a significant effect on a caller's decision to adopt CRBT—adoption is mostly shaped by adopters who are one-hop neighbors. When the group size is large, people tend to imitate indirect neighbors. When the group size is small, however, people tend to differentiate themselves from others.

We do not observe a significant effect of homophily from either direct or indirect friends—ρ_3 and ρ_4 are both insignificant across all networks. Such results imply that CRBT purchase is only affected by peer influence, but not homophily. The results are also consistent with some empirical studies in the literature. For example, Ma et al. (2015) showed that homophily does not affect individuals' purchase timing decision (when to buy); instead, it only affects product choice (what to buy). Since we treat all CRBT purchases as a binary decision and we do not consider product type, homophily not affecting the decision is expected.

Table 6. Meta-Analysis for Direct Influence (Cohesion), Smaller Subpopulations Group ($n \approx 200$) Pooled

		95% CI		
Subpopulation	Direct influence	Lower	Upper	Weight
1	0.070	0.026	0.11	1.1
2	0.060	0.026	0.14	0.15
3	0.065	0.0034	0.13	0.52
4	0.056	0.010	0.10	0.66
5	0.061	0.037	0.085	2.6

Summary effect = 0.063, 95% CI = (0.027, 0.099)

We then broke down our subpopulations to two groups based on size. One group of subpopulations is at $n \approx 200$, while the other group is at $n \approx 500$. Through the comparison of the two pooled mean effects, we can determine whether the role of direct and indirect peer influence varies with network size (H3A and H3B). The meta-analysis of the effect of direct peer influence in networks of $n \approx 200$ is shown in Table 6. The weights in the meta-analysis are determined by inverse variance.

In smaller networks ($n \approx 200$), direct peer influence has a pooled mean of 0.063, with a 95% confidence interval (0.027, 0.099) (Table 7). Larger networks ($n \approx 500$) have a pooled mean of 0.014, with a confidence interval (0.0069, 0.027). The results show that direct peer influence is stronger in smaller subpopulations ($n \approx 200$). The results of the meta-analysis confirm Hypothesis 3A that in smaller groups, individuals have stronger direct influence on each other than indirect influence.

For the five networks at $n \approx 200$, the summarized indirect influence is statistically significant with a size of −0.0067 and a 95% confidence interval (−0.011, −0.0025) (Table 8). This result confirms that for small sizes ($n \approx 200$), individuals tend to make different decisions from their two-hop neighbors in the network.

When the network size is $n \approx 500$, the summarized indirect influence effect is statistically significant at 0.0092, with a confidence interval (−0.011, −0.0025) (Table 9). The positive effect size suggests that in larger networks ($n \approx 500$), individuals tend to imitate their

Table 7. Meta-Analysis for Direct Influence, Larger Subpopulation Group ($n \approx 500$) Pooled

		95% CI		
Subpopulation	Direct influence	Lower	Upper	Weight
6	0.016	0.0063	0.026	0.81
7	0.013	0.0030	0.023	0.91
8	0.014	0.0048	0.023	1.0
9	0.016	0.0033	0.028	1.1
10	0.014	0.0090	0.019	1.2

Summary effect = 0.014, 95% CI = (0.0069, 0.027)

Table 8. Meta-Analysis for Indirect Influence, Smaller Subpopulations Group ($n \approx 200$) Pooled

Subpopulation	Direct influence	95% CI Lower	95% CI Upper	Weight
1	−0.0059	−0.0096	−0.0022	1.0
2	−0.0098	−0.014	−0.0059	0.82
3	−0.010	−0.017	−0.0034	0.90
4	−0.0079	−0.014	−0.0017	1.1
5	−0.0015	−0.0026	−0.00043	1.2
Summary effect = −0.0067, 95% CI = (−0.011, −0.0025)				

two-hop neighbors. Combining the results of Tables 7 and 8, we conclude that H3B is supported. Notably, the direction of indirect influence is different when the network size varies. This is a surprising new finding. This result can only be derived using subpopulations of different sizes and a model that supports multiple network autocorrelation terms.[8]

5.2. Robustness Checks

5.2.1. Propensity Score Matching. To further infer causality between the peer influences and CRBT purchase, we employed Difference-in-Differences (DID) matching strategies to explore the robustness of our hierarchical Bayesian model's estimation. The DID matching estimator compares the change of the purchase decision of individuals from the first time period to the last time period (12 weeks) by controlling for both direct and indirect peer influence. For each network size group, we constructed one control group and two treatment groups out of the population data (Table 10). The control group consists of individuals that are isolated from the phone call social network (represented as a Type 1 position in Table 10) from the first until the last time period, meaning that individuals in the control group are not affected by peer influence. The first treatment group consists of individuals who are isolated from the phone call network in the first time period but become a leaf node by the end of the last time period, connecting to exactly one individual from the network (represented as a Type 2 position). Individuals in the first treatment group are only affected by direct peer influence from their one and only neighbor who is a buyer. The second treatment group consists of individuals who are isolated from the phone call network in the first period but are connected in a triad with only one direct neighbor and one indirect neighbor (represented as a Type 3 position). Individuals in the second treatment group are affected by both direct and indirect peer influences. We focus on the network position types as the event of interest. Following the position type change, the purchase decisions of individuals in the two treatment groups are compared with those of matching individuals in the control group. The impact of direct influence is inferred by comparing the difference between the purchase probability change in the first treatment group and the control group. Individuals in the second treatment group are under a similar amount of direct influence as individuals in the first treatment group, because individuals in both groups have the same number of direct neighbors (one and only one). Given that the direct influence is controlled, the impact of indirect influence can be examined by comparing the difference of the purchase probability changes in the two treatment groups. The matching parameters include attributes representing homophily, such as age, gender, call duration, etc. Also, degree centrality is added in the matching variables to control for the peer influence of nonadopters. We employed propensity score matching (PSM) (Leuven and Sianesi 2003) for average treatment effects estimation. Finally, PSM is run on each smaller and larger network size group, respectively. The results are shown in Tables 10 and 11.

Table 9. Meta-Analysis for Indirect Influence, Larger Subpopulations Group ($n \approx 500$) Pooled

Subpopulation	Direct influence	95% CI Lower	95% CI Upper	Weight
6	0.0066	0.0043	0.0089	0.63
7	0.0098	0.0028	0.017	1.1
8	0.0010	0.0004	0.0016	1.1
9	0.034	0.019	0.049	0.97
10	0.0049	0.0032	0.0066	1.2
Summary effect = 0.0092, 95% CI = (0.0048, 0.014)				

Table 10. Network Position Type Change of Control and Treatment Groups Over Time

	Size	First time period (week 1)	Last time period (week 12)
Control group	2,000	Type 1	Type 1
Treatment 1	2,000	Type 1	Type 2
Treatment 2	2,000	Type 1	Type 3

Table 11. Average Treatment Effect of Peer Influences on Purchase of Individuals in Control and Treatment Groups Over Time

	PSM Smaller ($n \approx 200$)	PSM Larger ($n \approx 500$)
Treatment 1 vs. control	0.31* (0.14)	0.23* (0.10)
Treatment 2 vs. control	0.27* (0.11)	0.28* (0.13)
Treatment 2 vs. treatment 1	−0.070** (0.027)	0.061** (0.023)

Note. The data demonstrate the mean and standard deviation (in parentheses) of the average treatment effects.
$^{*}p < 0.01$; $^{**}p < 0.05$.

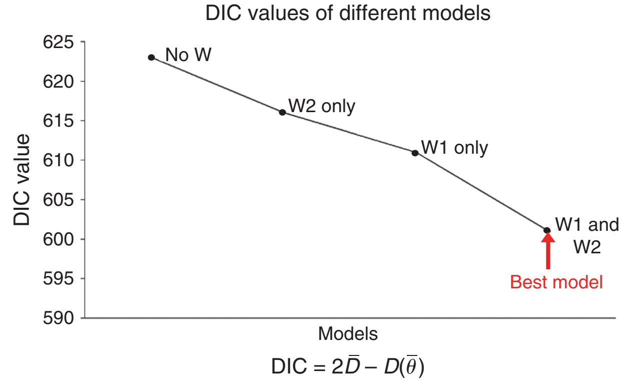
As we expected, the DID matching results indicate that the peer influences will significantly affect the CRBT purchase decision in both network size groups. In a smaller network size group ($n \approx 200$), when an individual is only affected by direct peer influence, she is 31% more likely to purchase CRBT ($p < 0.01$) than the matched player from the control group. When affected by both direct and indirect peer influences, an individual is 27% more likely to purchase CRBT ($p < 0.01$) than the matched sample from the control group. Such an increase is 7% less likely ($p < 0.05$) than for individuals in the direct influence only group. In a larger network size group ($n \approx 500$), when an individual is only affected by direct peer influence, she is 23% more likely to purchase CRBT ($p < 0.01$) than the control group. When affected by both direct and indirect peer influences, she is 28% more likely to purchase CRBT ($p < 0.01$). Such an increase is 6.1% more likely ($p < 0.05$) than for individuals in the direct influence only group. The PSM estimators confirm our finding that in a smaller network group, direct peer influence leads to a higher chance of CRBT purchase, while indirect influence has a cancellation effect—less of a chance among individuals. Meanwhile, in a larger network group, both direct and indirect influence leads to more CRBT purchases, and indirect influence has a reinforcement effect with regard to direct influence. All results confirm what we find in the main analysis.

5.2.2. Goodness-of-Fit Test. A significant network autocorrelation term does not necessary mean a model with two terms is better. We used the deviance information criterion (DIC) to measure our model's goodness of fit. It is similar to the Akaike information criterion (AIC) and Bayesian information criterion (BIC) model selection criteria, and we calculated from the likelihood of samples generated from Markov Chain Monte Carlo (MCMC) draws. The definition of DIC is shown below. Models with the lowest DIC value represent the best model

$$\bar{D} = E[-2\log(p(y \mid \theta))]$$
$$D(\bar{\theta}) = -2\log(p(y \mid \bar{\theta})).$$

As shown in Figure 5, when we have both peer influence terms (direct and indirect), the model is better. The model containing a term representing only direct peer influence is better that one with only indirect peer influence. The best model, the one with the lowest DIC, is one that includes both the direct and indirect influences simultaneously. Thus, we should include both network terms in our models. The goodness-of-fit test confirmed that CRBT adoption is influenced by both direct peer influence and indirect peer influence.

Figure 5. (Color online) Deviance Information Criterion (DIC) of Four Competing Models



5.2.3. Additional Robustness Checks. To confirm the robustness of peer influence on CRBT adoption, we also took the temporal information of both network connection (phone call conversation) and purchase decision into consideration. Since the entries of the two network terms $\mathbf{W}_1$ and $\mathbf{W}_2$ are defined by connections aggregated over the three-month period, we wanted to confirm whether the connections among users created in the earlier period still have the same pattern of (direct and indirect) influences on CRBT purchase in the latter period. Consider the model represented by the matrix form equations below

$$\mathbf{y}' = \mathbf{1}(\mathbf{z}' > 0),$$
$$\mathbf{z}' = \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\theta}' + \boldsymbol{\eta} + \boldsymbol{\varepsilon},$$
$$\boldsymbol{\theta}' = \rho_1\mathbf{W}_1''\boldsymbol{\theta}' + \rho_2\mathbf{W}_2''\boldsymbol{\theta}' + \rho_3\mathbf{W}_1''\mathbf{H}' + \rho_4\mathbf{W}_2''\mathbf{H}' + \mathbf{u}.$$

The model is the same as in Section 4, but with different values for the component terms. The vector $\mathbf{y}'$ represents observed CRBT purchases of all users in the second half of the three-month period; vector $\mathbf{z}'$ represents all people's latent preference in the same time period, resulting from an indicator function of value being larger than zero; matrix $\mathbf{X}'$ represents all of the users' attributes in the same time period $\mathbf{y}'$; people's interdependent component of preference, $\boldsymbol{\theta}'$, can be represented as the sum of the autocorrelation terms $\rho_1\mathbf{W}_1''\boldsymbol{\theta}' + \rho_2\mathbf{W}_2''\boldsymbol{\theta}' + \rho_3\mathbf{W}_3''\mathbf{H}' + +\rho_4\mathbf{W}_4''\mathbf{H}'$. The $\mathbf{W}_1''$ and $\mathbf{W}_2''$ terms are derived from the network connections created in the first half of the three-month period. The homophily matrix $\mathbf{H}'$ is also defined using the attributes in the second half of the three-month period. If the estimates are consistent with those in Table 5, then we can confirm the robustness of our results. The estimated results are shown in Table 12.

All results are very similar to those in Table 5 using aggregated CRBT purchases and network structure. The robustness of the model is thus further supported.

We then recoded $\mathbf{y}'$ as the vector representing observed CRBT purchase decisions of all individuals

Table 12. Robustness Check of Analysis Using the mNAP Model, Direct vs. Indirect Influence (CRBT Purchase from the Second Half of the Three-Month Period, Network from the First Half of the Three-Month Period)

	Subpopulations									
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Degree centrality	0.013^{*}	0.036^{*}	0.031^{*}	0.0045	0.068^{*}	0.038	0.046^{*}	0.020	0.040^{*}	0.027^{*}
	(0.0096)	(0.011)	(0.011)	(0.020)	(0.017)	(0.024)	(0.021)	(0.033)	(0.0043)	(0.0075)
Direct influence	0.074^{*}	0.063^{*}	0.060^{**}	0.059^{**}	0.056^{*}	0.050^{**}	0.045^{*}	0.045^{*}	0.052^{**}	0.040^{*}
	(0.023)	(0.021)	(0.026)	(0.027)	(0.020)	(0.022)	(0.013)	(0.016)	(0.023)	(0.015)
Indirect influence	-0.0061^{**}	-0.010^{**}	-0.0090^{**}	-0.0090^{**}	-0.0017^{***}	0.0064^{**}	0.011^{**}	0.0011^{**}	0.033^{*}	0.0046^{**}
	(0.0030)	(0.0040)	(0.0040)	(0.0043)	(0.00089)	(0.0031)	(0.0048)	(0.00058)	(0.013)	(0.0023)
Direct homophily	0.0037	0.0037	0.0011	0.0018	0.0012	0.042	0.0016	0.0017	0.00078	0.0042
	(0.0036)	(0.0038)	(0.0033)	(0.0025)	(0.0041)	(0.071)	(0.0020)	(0.0021)	(0.0037)	(0.087)
Indirect homophily	0.00026	0.00035	9.6×10^{-5}	0.00012	0.00011	0.0029	0.00015	0.00014	6.8×10^{-5}	0.00029
	(0.0036)	(0.0028)	(0.00031)	(0.00018)	(0.00031)	(0.0056)	(0.00018)	(0.00021)	(0.00027)	(0.0067)

$^{*}p < 0.01$; $^{**}p < 0.05$; $^{***}p < 0.10$.

in the second and third months of the three-month period, and we derived the $\mathbf{W}_1''$ and $\mathbf{W}_2''$ terms from the network connection created in the first month. The definition of the rest of the model is the same as in the paragraph above. The estimation results are presented in Table C1 (Online Appendix C). Those results are still consistent with the results shown in Table 12.

We also changed the window width to redefine $\mathbf{y}', \mathbf{W}_1''$, and $\mathbf{W}_2''$ terms. Using the CRBT purchase decisions of all users in the third month to define $\mathbf{y}'$ and deriving the $\mathbf{W}_1''$ and $\mathbf{W}_2''$ terms from the network connection created in the first and second months, the rest of the model keeps the same structure as the previous two robustness checks. The estimation results are presented in Table C2 (Online Appendix C). The results are consistent with those in both Table 12 and Table C1 (Online Appendix C).

To confirm whether peer influence has the same pattern of impact on the number of CRBT purchases as a binary purchase decision, we also ran the same data set on a model supporting a continuous type response variable. The model is an extension of Doreian's (1989) two regimes of network effects autocorrelation model with an additional third network term. The model is defined using the matrix notation shown below

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho_1\mathbf{W}_1\mathbf{y} + \rho_2\mathbf{W}_2\mathbf{y} + \rho_3\mathbf{W}_1\mathbf{H} + \rho_4\mathbf{W}_2\mathbf{H} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is the vector of the observed number of CRBT purchases for all callers included in the data. Both $\mathbf{X}$ and $\boldsymbol{\beta}$ are the same as those in the model defined in Section 4. Both network terms $\mathbf{W}_1$ and $\mathbf{W}_2$ are the same as those defined in Section 4. Matrix $\mathbf{H}$ still represents measures describing homophily. Entries of these three matrices were also derived from connections created over the three-month period. The autocorrelated response variable $\mathbf{y}$ is also on the right-hand side. The product term $\mathbf{W}_i\mathbf{y}$ represents the aggregated number of CRBT purchases from all neighbors in network i for each of the focal users. ρ_i, $i = \{1, 2\}$ is the corresponding coefficient for term $\mathbf{W}_i\mathbf{y}$. The product term $\mathbf{W}_i\mathbf{H}$, $i = \{3, 4\}$ represents the aggregated impact of homophily from all neighbors in network i for each of the focal users. ρ_i, $i = \{3, 4\}$ represents the corresponding coefficients. ε is the random error term. Estimation results are shown in Table C3 in Online Appendix C, and they have the same pattern as all of the previous models. This again confirms the robustness of our results.

Similar to the tests performed for binary CRBT purchase outcomes, we took the temporal information of both the network connection and purchase time into consideration, and we redefined the model as follows:

$$\mathbf{y}' = \mathbf{X}'\boldsymbol{\beta} + \rho_1\mathbf{W}_1''\mathbf{y}' + \rho_2\mathbf{W}_2''\mathbf{y}' + \rho_3\mathbf{W}_1''\mathbf{H}' + \rho_4\mathbf{W}_2''\mathbf{H}' + \boldsymbol{\varepsilon},$$

where $\mathbf{y}'$ is the vector representing the observed number of CRBT purchases by all individuals in the second half of the three-month period; $\mathbf{X}'$ is the matrix representing all of the individuals' attributes in the same time period as $\mathbf{y}'$; $\mathbf{H}'$ is the matrix representing homophily dimensions in the same time period; the $\mathbf{W}_1''$ and $\mathbf{W}_2''$ terms representing network structures are derived from the conversation relationships that occurred in the first half of the three-month period. Results are shown in Table C4 (Online Appendix C), and they have a similar pattern as the results in Table 12.

We then used the number of all users' CRBT purchases in the second and third months to define $\mathbf{y}'$, the attributes accounting for homophily in the same time period to define $\mathbf{H}'$, and the network connection created in the first month of the three-month period to define the $\mathbf{W}_1''$ and $\mathbf{W}_2''$ terms. The rest of the model kept the same structure as the previous two robustness tests. The estimation results are presented in Table C5 (Online Appendix C). The results are consistent with Table 12. The test of using the number of CRBT purchases of all users in the third month as $\mathbf{y}'$, and the network connection created in the first two months of the three-month period to define the $\mathbf{W}_1''$ and $\mathbf{W}_2''$ terms is also done. The estimation results are shown in Table C6 (Online Appendix C), and they are also consistent with the results in Tables C3–C5 (Online Appendix C). All robustness checks show that the results with regard to direct and indirect peer influence, along with centrality and homophily, are robust. Taken together, CRBT adoption is affected by all three types of peer influence (direct, indirect, and centrality), but *not* homophily, extending the work of Aral et al. (2009). These results imply that for a decision that is mainly affected by peer influence, such as the case of CRBT, the effect of homophily could be less influential than direct and indirect peer influence and centrality.

6. Discussion

6.1. Key Findings

Using the subpopulations extracted, we analyzed the effects of direct and indirect peer influence on the adoption of CRBT using an extended mNAP model with new implementation to support the complexity of large-scale social networks. Our study is one of very few to investigate several peer influences on technology diffusion as a binary variable in social networks. Our results show that the adoption of CRBT is consistently predicted by direct peer influence. When the size of a subpopulation is small (about 200), CRBT adoption is negatively affected by indirect peer influence (structural equivalence); when the size of a subpopulation is large (about 500), adoption is positively affected by indirect peer influence. Between these two network effects, direct peer influence has a statistically significant larger coefficient size than indirect peer influence. Such a result is obtained when other peer influences—centrality and homophily—are controlled for, so its robustness is assured. Based on these results, businesses should use particular strategies for networks of different sizes. If businesses want to trigger higher adoption rates, then for smaller groups, they only need to focus on individuals with many direct connections, while for larger groups, they should not only focus on individuals who are popular but also those who have many common one-hop friends with two-hop friends.

6.2. Contributions and Implications for Theory and Methodology

The first and foremost implication of our research is the reconciliation of a classic yet inconclusive theoretical problem: identifying which peer influence affects technology diffusion. The long-held debate about the relative impact of two popular network models—cohesion (direct influence) versus structural equivalence (indirect influence)—on technology diffusion in social networks still persists. The literature has assumed that only a single peer influence affects diffusion: some researchers still believe that only direct peer influence triggers technology diffusion, while others believe that indirect peer influence from two-hop neighbors matters more, and both camps have found empirical support for their theoretical claims (e.g., Coleman 1958, Burt 1987, Leenders 2002, Van den Bulte and Lilien 2001). One gap in the literature is that other than Coleman's classical medical innovation data, few new data sets have been used to address this important question. Reconciling these findings is very important, because a social network is a key medium of technology diffusion, and figuring out which social network effect drives this diffusion can help us decide what targeting strategies to use to encourage it.

By using a new statistical model, the proposed enhanced mNAP that supports panel data, we are able not only to show which peer influence is more influential but also to capture when both peer influences are significant, even in opposite directions (different signs). Controlling for homophily using exogenous personal attributes and centrality as a covariate assures the robustness of our results. Without our new implementation of the mNAP model, which supports three network autocorrelation terms and other covariates (for centrality and homophily to be controlled for), such results could not be properly identified. Our analyses, using multiple network sizes, provide new insights into technology diffusion since the size of peer influence and its direction are associated with network size. Extraction of multiple subpopulations demonstrates that in a large-scale social network, subnetworks with high internal density may not be big, and the variations of group size may not be wide. In our cellular phone caller-callee (CRBT) network, network size only comes in two levels—about 200 and 500. This finding implies that densely connected subnetworks as subpopulations when analyzing large social networks can be a feasible and valid strategy. Our study is a modest step toward tackling large-scale social network data with multiple network terms. With the emergence of social media and social communities, increasingly more social network datasets are large scale, and many types of relationships and patterns can be observed in network data. Our models should thus enable researchers to solve a variety of problems in such large-scale network data.

In the context of social networks with millions or even billions of nodes, there is high heterogeneity among different subnetworks. Such data may be intractable because of their large size. To resolve these problems, we constructed a model to make large network data be less heterogeneous and tractable, and we extended a recent statistical model to estimate the two types of peer influence. We also used an innovative algorithm, T-CLAP, to extract subpopulations from the population network. This method not only concentrates the analysis on individual networks that are relatively self-contained, but it also preserves the properties and structure of the network. Also, an analysis of influence maximization in online settings should consider the clusters of peer influence and the subnetworks within which individual content creators are embedded. Focusing on an individual content creator might overstate or understate the effect of peer influence.

Our study offers guidance on using subnetwork extraction strategy to social media analytics in networks. Since we examined technology diffusion through social network connections (caller-callee relationships), our study implies that it is necessary to consider the degree of interaction across individuals to analyze technology diffusion in a group of networked individuals. The methodological issues involved in identifying individuals' probability for technology adoption could also apply to other settings beyond phone networks. Since many social media phenomena are of much interest to practitioners, such as app adoption on iTunes and Facebook (Aral and Walker 2011) and content consumption on YouTube (Susarla et al. 2012), all large networks, this method could find broader applications. Our ability to identify endogenous peer influences on individual behavior while controlling for homophily can be generalized to other studies of social influence in networks.

Peer influence could be widely applied to many IS phenomena, such as adoption of apps in online social media networks, content creation on social media platforms, and software adoption in a network. On social network sites, people are able to observe the apps adopted by their direct friends; at the same time, they read about the apps adopted by indirect friends through their direct friends. Indirect influence may be even stronger than direct influence, because an individual may want to impress her friends and make them write posts about her. Indirect peer influence may play an important role in software adoption, since a user can adopt software because of indirect peer influence, even if her direct friends did not adopt the software. Given that many products, such as mobile apps, music, and videos, are on networked platforms, our results can be extended to other IS phenomena, such as apps and software adoption. Our framework for processing and analyzing network data enables us to generalize to other social networks. With access to social media data from social networks, many IS researchers have large-scale network data, although there are still no straightforward approaches for analyzing large-scale network data.

6.3. Contributions and Implications for Practice

The key to the growth of CRBT adoption on cellular phone platforms is to target existing adopters in social groups and let them use peer influence to affect neighbors in a social network and simultaneously promote the product to individuals who have similar personal attributes as the adopters. The penetration rate of CRBT in Asia, Europe, and Africa indicates that there is still space for CRBT market growth. In 2011, Verizon added CRBT to its Music V CAST with the purpose of making CRBT adoption more convenient and potentially encouraging CRBT adoption. However, our results show that CRBT adoption is not significantly associated with cellular phone users' similarity in personal attributes (homophily in terms of demographics, behavior, social circle, and location), but rather with direct and indirect peer influence. This is consistent with existing empirical studies, such as Ma et al. (2015)—homophily does not affect individuals' purchase decision timing (when to buy); instead, it only affects product choice decisions (what to buy). We treat all purchases of CRBT as a binary decision, and we do not consider product type; thus, our results are still consistent with those of Ma et al. (2015). Cellular phone service providers should thus invest their marketing efforts into leveraging adopters to affect their nonadopter neighbors, both direct and indirect.

Our results show that networks of various sizes have different indirect influence directions and signs. Indirect influence is positive and significant in larger networks ($n \approx 500$), but negative and significant in smaller networks ($n \approx 200$). Thus, service providers should design different target marketing schemes for larger versus smaller networks. For smaller networks, service providers should target adopters with many direct neighbors who can affect their one-hop neighbors through direct ties. Service providers should also avoid adopters with many indirect (two-hop) friends, because these friends are *not* likely to follow what their indirect adopter friends do. For larger social networks, service providers should target adopters with many direct and indirect friends, because adopters could affect both types of friends through direct and indirect ties, and both types of peer influence will increase their friends' probability of adoption.

6.4. Limitations and Suggestions for Future Research

There are some limitations to our research that must be addressed. First, we treat social network terms, direct influence, indirect influence, and homophily, as fixed effects, as a model containing these terms as random effects has not been implemented yet. Future research can develop a new model that takes network autocorrelation terms as random effects. Second, our data does not contain many personal attributes of the cellular phone subscribers other than age and gender. Based on the empirical results using age and gender, there is no strong evidence that indicates that the network connections are associated with these attributes.

Still, it is possible that the creation of network connections is affected by other personal attributes. Third, other possible shocks on CRBT adoption, for example, a marketing campaign, are not included in our data. Fourth, although we explicitly controlled for measures of homophily in our model to empirically show that with several dimensions controlled for, the effect of homophily is not empirically significant, there is still a possibility that some unobserved variables that are correlated with homophily are not captured in our data. Fifth, membership in smaller and larger groups is possibly endogenous. Individuals in smaller groups, on average, are different in attributes that are not included in our data set, from those who are in larger groups. Thus, they behave differently with respect to differentiating themselves with respect to indirect neighbors.

Future research could also explore the effect of weak ties (asymmetric phone calls). Current research has not explored whether peer influence among individuals connected by weak ties exists. An empirical study on whether weak ties affect diffusion can help businesses determine whether asymmetric relationships need to be considered when designing a conversion strategy, as well as its associated cost. For example, if peer influence size among individuals connected by weak ties is not much less than that among individuals with strong ties, and more people are connected with weak ties, then the marketing scheme should still target individuals connected with weak ties, because potentially more people could become adopters. We can also explore the case where individuals' latent preferences, as well as these individuals' observed decisions, are affected by their peers. Existing network autocorrelation models only consider situations where peer influence works as an interdependent latent preference, regardless of observing neighbors' decisions, or imitates the neighbor's decision only when the decision is observed. If peer influence only happens when individuals observe what their neighbors do, businesses should make their marketing targets know that their neighbors have adopted the focal technology or product. To infer the causality inference between (direct and indirect) peer influence and CRBT adoption, one should take both the time of network connection creation and the time of adoption into consideration. If adoption happens after a new relationship is created between an adopter and a nonadopter, the causality of peer influence on CRBT adoption can be strengthened further.

More in-depth investigation of the direction of indirect influence in different sizes of networks is needed. We empirically found the opposite direction of indirect influence in smaller and larger networks, respectively. Our initial results suggest that individuals have divergent ways of demonstrating their personality in different sizes of networks. In smaller networks, personality is associated with different purchase decisions. However, in larger networks, personality is associated with the genre of CRBT. Future research could include additional subnetworks of different sizes to study whether the pattern of individuality in indirect influence is persistent. Considering that network size may be associated with individual attributes, future research could also include more attributes in the aforementioned networks.

Finally, our meta-analyses of multiple subpopulations return a single value for the population moments parameters. The assumption is that all subpopulations are homogenous, and their parameters are the same as the population parameter. However, in the focal context of social networks, when a network gets larger (especially > 1 million), it is very likely that subnetworks are heterogeneous and can be classified in multiple groups. In this case, one single number cannot represent the parameter of all subnetwork groups. Specifically, a distribution of a hyperparameter for the population would be a better solution. Accordingly, future research could design a new sophisticated Bayesian model to estimate the distribution of population parameters.

7. Conclusion

Technology diffusion has long been recognized as a very important challenge for IS research. In a very well-connected world, through cellular phones, social media, and other information technologies, people are influenced by peers when making decisions. Two key types of peer influence were identified in the literature and are believed to shape adoption decisions: direct influence and indirect influence. Direct influence comes from focal individuals' one-hop neighbors, that is, when an individual has many one-hop neighbors who are adopters, she is likely to adopt. Indirect peer influence comes from an individual's two-hop neighbors who share common one-hop neighbors. It is believed that the more common friends two individuals share, the more similar social positions these two individuals have and therefore make the same adoption decisions. Thus, peer influence can be considered as coming from two-hop neighbors, that is, when an individual has many two-hop neighbors with whom the focal friend shares common friends who are adopters, she is likely to adopt. Both direct and indirect peer influence can explain the phenomenon that people who are embedded in social networks tend to make similar decisions. However, the debate about which type of peer influence dominates has not been settled in the literature, because these two types of peer influence lead to different conversion schemes. If direct peer influence is the driver of adoption decisions, then businesses should target adopters with many one-hop neighbors, because adopters can affect their immediate neighbors to also adopt. Yet if indirect peer influence is the driver, then businesses should target adopters who share the same group of common one-hop friends with many two-hop friends, because the adoption decision will happen through the two-hop relationship. Comparing these two types of peer influence is challenging, especially in large-scale social networks. First, such study requires a binary choice network autocorrelation model that supports multiple types of peer influence. Second, since both social influences and homophily could be the reason for technology adoption, a model that supports three network terms—direct influence, indirect influence, and homophily, respectively—with homophily as a control variable is required. However, such a model has not been implemented nor has it been empirically tested in the literature.

Moreover, the size of the social network presents computation and heterogeneity challenges. To address these practical challenges, using a large-scale social network data set from cellular phone call conversations, we examined the role of direct and indirect peer influence in CRBT adoption, while controlling for network centrality and homophily. Our analysis is built on a framework that allows us to process large-scale network data that contains a process to extract multiple small network subpopulations and a meta-analysis for summarizing the results from multiple tests. Our study thus contributes to the literature by providing empirical evidence to settle the long-held debate in the literature about which type of peer influence is the main driver of technology diffusion.

The proposed framework in this study also makes contributions to methodology, because we implemented the first hierarchical Bayesian model to accommodate three network autocorrelation terms (accounting for direct and indirect peer influence), while including homophily as a control (network) term. We find that in real-life, cellular phone conversation networks, people tend to form groups of only two sizes—a smaller size with about 200 people and a larger size with about 500 people. We further find that both direct and indirect peer influences have a strong effect on CRBT adoption in both small and large networks. Although the pattern of direct peer influence is consistent in both groups, the pattern of indirect influence is not. People only make similar decisions to those of their two-hop neighbors when the group size is large (about 500). However, when the group is smaller, the pattern goes in the other direction—people do not make the same decision as their two-hop neighbors. Our explanation is that when the network is smaller, people show their individuality by making a distinctive adoption decision (that is, not purchasing CRBT), while when the network is larger, people demonstrate their individuality by purchasing CRBT of a different music genre. If an indirect neighbor purchases CRBT in one genre, for example, rock, the focal individual also purchases CRBT, but the CRBT belongs to a different music style, for example, hip-hop. Our findings show that there are two different strategies of showing individuality in smaller and larger networks, respectively.

Our study is one of the very few to examine both network peer influence terms on diffusion, especially in large social networks. The results are interesting—both direct and indirect influences affect diffusion—and notably in the opposite direction. The identification of peer influence is also empirically supported. The findings of our work can provide evidence for reconciling these two types of peer influence in large-scale networks. We have also controlled for homophily, although it is not significant in the adoption of CRBT. Our model thus provides a means to reconcile the role of both peer influences and homophily on adoption. Such reconciliation was considered a very difficult challenge previously because of methodological limitations. Given our methodological advances, future research could apply our model to other adoption phenomena in social networks to examine whether the patterns observed in this study are consistent with other contexts.

Finally, the results of our study have direct implications for practitioners—they can advise managers on designing targeted strategies to encourage more people to adopt CRBT and other networked products. Specifically, businesses could design different strategies for smaller versus larger networks of consumers, thereby stressing the importance of network size in affecting product adoption and technology diffusion. For smaller networks, businesses should design marketing schemes targeting opinion leaders who can exert peer influence on many direct neighbors (but only on a few indirect neighbors), whereas for larger networks, businesses should target opinion leaders who can exert peer influence to both direct and indirect neighbors.

Acknowledgments

The authors would like to thank the senior editor and the associate editor for their support and guidance. The authors would also like to thank the anonymous reviewers for their constructive comments and suggestions throughout the review process.

Endnotes

[1] Direct ties with one-hop neighbors, formally defined as *cohesion*.

[2] Indirect ties with two-hop neighbors, formally defined as *structural equivalence*.

[3] CRBT replaces plain ringback tones with music for a caller to hear as she waits for the callee to answer. CRBT is becoming one of the most attractive mobile contents, with estimated revenues of $235 million in the United States in 2009 (Broadcast Music Inc. 2009). The penetration rate of CRBT is even larger in Asia and Africa; for example, 95% of the digital music market in Indonesia comes from CRBT. Unlike other social networks, which are often extracted from social

media websites, the interaction between a caller-callee pair entails a stronger intent-to-communicate, implying that a CRBT network is a good approximation of a social network. Thus, it is useful to understand how diffusion unfolds in a CRBT social network: whether new adopters adopt CRBT because of direct influence from existing adopters (one-hop neighbors), or because of indirect influence from two-hop neighbors. Despite the nonexclusive effects of these two peer influences, we simultaneously examine their effects on CRBT diffusion in social networks. CRBT adoption is represented by a binary variable within the large-scale cellular phone communication network.

[4] Centrality only takes the number of the direct neighbors into account without considering the neighbors' adoption decisions, while direct influence takes both the number of direct neighbors and their decisions into account.

[5] As the variation in the number of one-hop neighbors could be very large for all individuals, we use the ratio of adopters among neighbors instead of the absolute number of adopters; this is because one single neighbor's influence could be different for an individual with a small number of neighbors from another individual with a large number of neighbors.

[6] In organizations, individuals who are structurally equivalent typically have the same role in the organization, and thus the model is also called the "role equivalence" model. An individual is exactly structurally equivalent to her neighbor if both of them share the same set of neighbors.

[7] Our mNAP model is estimated using MCMC. The detailed estimation steps are shown in Online Appendix A.

[8] We should point out that the change from indirect to direct neighbor is trivial. First, the average proportion of the addition or the removal of a friend relationship from one-time period to the next is less than 5%. Second, once an indirect neighbor becomes a direct neighbor, she is considered as one who poses a direct influence to the focal node, and she is thus no longer considered as an indirect peer influence.

References

Agarwal R, Gupta AK, Kraut R (2008) Editorial overview—The interplay between digital and social networks. *Inform. Systems Res.* 19(3):243–252.

Ahuja G (2000) Collaboration networks, structural holes, and innovation: A longitudinal study. *Admin. Sci. Quart.* 45(3):425–455.

Allenby GM, Rossi PE (1998) Marketing models of consumer heterogeneity. *J. Econometrics* 89(1–2):57–78.

Aral S, Walker D (2011) Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Sci.* 57(9):1623–1639.

Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci.* 106(51):21544–21549.

Bass F (1969) A new product growth for model consumer durables. *Management Sci.* 15(5):215–227.

Bernheim BD (1994) A theory of conformity. *J. Political Econom.* 102(5):841–877.

Bonacich P (1987) Power and centrality: A family of measures. *Amer. J. Sociol.* 92(5):1170–1182.

Bott H (1928) Observation of play activities in a nursery school. *Genetic Psych. Monographs* 4(1):44–88.

Bramoullé Y, Djebbari H, Fortin B (2009) Identification of peer effects through social networks. *J. Econometrics* 150(1):41–55.

Brancheau CJ, Wetherbe CJ (1990) The adoption of spreadsheet software: Testing innovation diffusion theory in the context of end-user computing. *Inform. Systems Res.* 1(2):115–143.

Broadcast Music Inc. (2009) BMI announces 2009 mobile music market projections. https://www.bmi.com/press/entry/538235/.

Burkhardt ME (1994) Social interaction effects following a technological change: A longitudinal investigation. *Acad. Management J.* 37(4):869–898.

Burt RS (1982) *Toward a Structural Theory of Action: Network Models of Social Structure, Perception, and Action*, Quant. Stud. Soc. Relations (Academic Press, New York).

Burt RS (1987) Social contagion and innovation: Cohesion versus structural equivalence. *Amer. J. Sociol.* 92(6):1287–1335.

Chatterjee R, Eliashberg J (1990) The innovation diffusion process in a heterogeneous population: A micromodeling approach. *Management Sci.* 36(9):1057–1079.

Coleman JS (1958) Relational analysis: The study of social organization with survey methods. *Human Organ.* 17(4):28–36.

Coleman JS, Katz E, Menzel H (1966) *Medical Innovation: A Diffusion Study* (Bobbs-Merrill, Indianapolis).

Doreian P (1982) Maximum likelihood methods for linear models: Spatial effects and spatial disturbance terms. *Sociol. Methods Res.* 10(3):243–269.

Doreian P (1989) Two regimes of network effects autocorrelation. Kochen M, ed. *The Small World* (Ablex Publishing, Norwood, NJ).

Dunbar R (1992) Neocortex size as a constraint on group size in primates. *J. Human Evolution* 20:469–493.

Duncan OD, Haller AO, Portes A (1968) Peer influences on aspirations: A reinterpretation. *Amer. J. Sociol.* 74(2):119–137.

Fang X, Hu P, Li Z, Tsai W (2013) Predicting adoption probabilities in social networks. *Inform. Systems Res.* 24(1):128–145.

Fujimoto K, Valente TW (2012) Social network influences on adolescent substance use: Disentangling structural equivalence from cohesion. *Soc. Sci. Medicine* 74(12):1952–1960.

Goolsbee A, Klenow PJ (2002) Evidence on learning and network externalities in the diffusion of home computers. *J. Law Econom.* 45:317–343.

Hanneke S, Fu W, Xing EP (2010) Discrete temporal models of social networks. *Electronic J. Statist.* 4:585–605.

Hanneman R, Riddle M (2005) Introduction to social network methods. University of California Riverside, http://www.faculty.ucr.edu/~hanneman/nettext/.

Harkola J, Greve A (1995) Diffusion of technology: Cohesion or structural equivalence? *Acad. Management Proc.*, 422–426.

Kamakura WA, Russell GJ (1989) A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* 26(4):379–390.

Katona Z, Zubcsek PP, Sarvary M (2011) Network effects and personal influences: The diffusion of an online social network. *J. Marketing Res.* 48(3):425–443.

Krackhardt D, Stern RN (1988) Informal networks and organizational crises: An experimental simulation. *Soc. Psych. Quart.* 51(2): 123–140.

Lazarsfeld PF, Merton RK (1954) Friendship as a social process: A substantive and methodological analysis. Abel T, Page CH, eds. *Freedom and Control in Modern Society* (Van Nostrand, New York), 18–66.

Leenders RT (1997) Longitudinal behavior of network structure and actor attributes: Modeling interdependence of contagion and selection. Doreian P, Stokman F, eds. *Evolution of Social Networks* (Routledge, Amsterdam), 165–184.

Leenders RT (2002) Modeling social influence through network autocorrelation: Constructing the weight matrix. *Soc. Networks* 24: 21–47.

Leuven E, Sianesi B (2003) PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components S432001, Department of Economics, Boston College, Chestnut Hill, MA.

Ma L, Krishnan R, Montgomery AL (2015) Latent homophily or social influence? An empirical analysis of purchase within a social network. *Management Sci.* 61(2):454–473.

Manski CF (1993) Identification of endogenous social effects: The reflection problem. *Rev. Econom. Stud.* 60(3):531–542.

Manski CF (2000) Economic analysis of social interactions. *J. Econom. Perspect.* 14(3):115–136.

McPherson JM, Smith-Lovin L (1987) Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *Amer. Sociol. Rev.* 52(3):370–379.

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual Rev. Sociol.* 27: 415–444.

Nair HS, Manchanda P, Bhatia T (2010) Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *J. Marketing Res.* 47(5):883–895.

Oinas-Kukkonen H, Lyytinen K, Yoo Y (2010) Social networks and information systems: Ongoing and future research streams. *J. Assoc. Inform. Systems* 11(2):62–67.

Ord K (1975) Estimation methods for models of spatial interaction. *J. Amer. Statist. Assoc.* 70(349):120–126.

Premkumar GRK, Nilakanta S (1994) Implementation of electronic data interchange: An innovation diffusion perspective. *J. Management Inform. Systems* 11(2):157–186.

Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph models for social networks. *Soc. Networks* 29(2):173–191.

Rogers EM (1962) *Diffusion of Innovations* (Free Press, New York).

Rogers EM, Kincaid LD (1981) *Communication Networks: Toward a New Paradigm for Research* (Free Press, New York).

Salton G (1989) *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer* (Addison-Wesley, Reading, MA).

Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociol. Methods Res.* 40(2):211–239.

Shriver SK, Nair HS, Hofstetter R (2013) Social ties and user-generated content: Evidence from an online social network. *Management Sci.* 59(6):1425–1443.

Smith TE, LeSage JP (2004) A Bayesian probit model with spatial dependencies. Pace KR, LeSage JP, eds. *Advances in Econometrics*, Spatial Spatiotemporal Econometrics, Vol. 18 (Elsevier, Oxford, UK), 127–160.

Snijders TAB (2001) The statistical evaluation of social network dynamics. Sobel ME, Becker MP, eds. *Sociological Methodology*—2001, Vol. 31 (Basil Blackwell, Boston), 361–395.

Snijders TAB, van Duijn MAJ (1997) Simulation for statistical inference in dynamic network models. Conte R, Hegselmann R, Terna P, eds. *Simulating Social Phenomena* (Springer, Berlin), 493–512.

Snijders TAB, Koskinen JH, Schweinberger M (2010) Maximum likelihood estimation for social network dynamics. *Ann. Appl. Statist.* 4:567–588.

Steglich C, Snijders TAB, West P (2006) Applying SIENA: An illustrative analysis of the coevolution of adolescents' friendship networks, taste in music, and alcohol consumption. *Methodology* 2(1):48–56.

Strang D, Tuma NB (1993) Spatial and temporal heterogeneity in diffusion. *Amer. J. Sociol.* 99(3):614–639.

Susarla A, Oh JH, Tan Y (2012) Social networks and the diffusion of user-generated content: Evidence from YouTube. *Inform. Systems Res.* 23(1):23–41.

Trusov M, Bodapati AV, Bucklin RE (2010) Determining influential users in Internet social networks. *J. Marketing Res.* 47(4):643–658.

Valente TW (2005) Models and methods for innovation diffusion. Carrington PJ, Scott J, Wasserman S, eds. *Models and Methods in Social Network Analysis* (Cambridge University Press, Cambridge, UK), 98–116.

Van den Bulte C, Lilien GL (2001) Medical innovation revisited: Social contagion versus marketing effort. *Amer. J. Sociol.* 106(5): 1409–1435.

Yang S, Allenby GM (2003) Modeling interdependent consumer preferences. *J. Marketing Res.* 40(3):282–294.

Zhang B, Krackhardt D, Krishnan R, Doreian P (2011) An effective and efficient subpopulation extraction method in large social networks. *ICIS 2011 Proc. 7.* https://aisel.aisnet.org/icis2011/proceedings/breakthroughideas/7.

Zhang B, Thomas AC, Doreian P, Krackhardt D, Krishnan R (2013) Contrasting multiple social network autocorrelations for binary outcomes: With applications to technology adoption. *ACM Trans. Management Inform. Systems* 3(4):1–21.